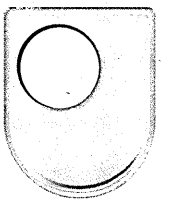
Establishment of Zygotic Transcription and Chromatin Organization in the Early *Drosophila* Embryo

Thesis

STOWERS INSTITUTE®
FOR MEDICAL RESEARCH



The Open University

Establishment of Zygotic Transcription and Chromatin Organization in the early *Drosophila* Embryo

A Thesis Submitted for the Degree of
Doctor of Philosophy
by

Kai Chen

The Stowers Institute for Medical Research,
an Affiliated Research Centre of the Open University

Kansas City, January 2013

DATE OF SUBMISSION : 8 JANUARY 2013
DATE of AWARD : 29 MAY 2013

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

To my lovely mom and daughter

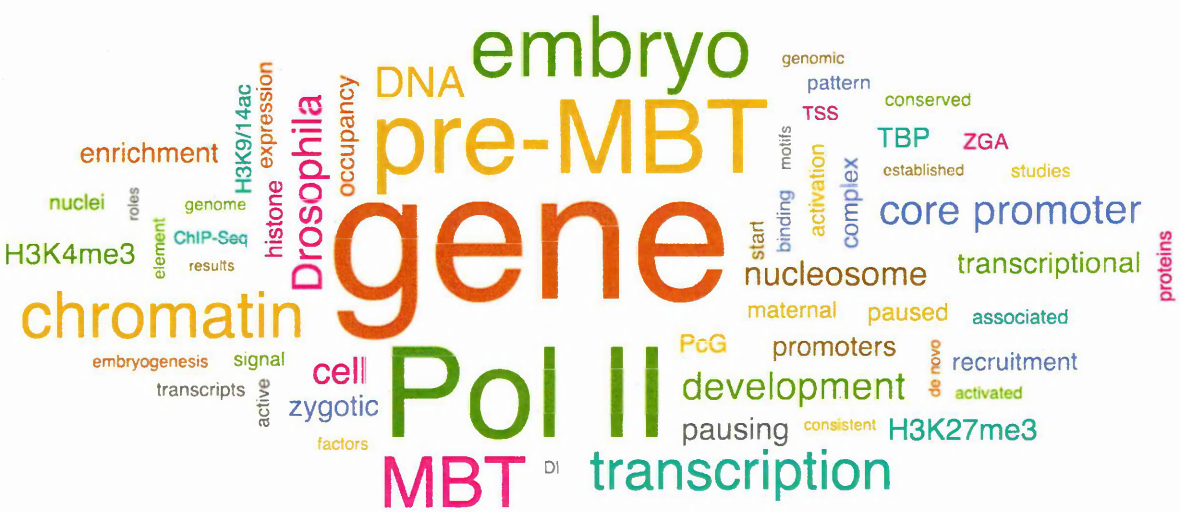
Acknowledgments

Working on this thesis was like a marathon, and I would not have been able to complete this memorable journey without the support of countless people over the past five years. First of all, I would like to especially thank my advisor, Dr. Julia Zeitlinger, not only for allowing me to work on this challenging project that became my dissertation, but also for her forbearance and understanding during the difficult times while I was developing protocols and writing my thesis. Her constant encouragement and invaluable suggestions made this work successful. I learned from her how to be a good scientist, which has always been my dream career.

I am deeply indebted to my committee members Dr. Joan Conaway, Dr. Robb Krumlauf, Dr. Susan Abmayr, Dr. Ting Xie, and my third party monitor, Dr. Arcady Mushegian, for their cordial suggestions and encouragement, which made it possible for me to finish this story. I am grateful to Dr. Leanne Wiedemann and Shelly Hornbuckle for coordinating my Open University committee meetings and for keeping track of the thesis timeline. I would also like to express my gratitude to Dr. Julie Ahringer and Dr. Marco Blanchette for being my external and internal examiners in reviewing this work.

I also appreciate the contributions of Jeff Johnston and Sam Meier for data processing and analysis, Wanqing Shao, Nina Keonecke, and Lydia Sun (Rushlow lab) for contributing ChIP-Seq experiments, Cindi Staber for all kinds of lab support, Axel Shum, Dr. Malini Natajaran, and Holly Zink for proofreading my drafts of this thesis. I also appreciate all the other lovely members of the Z lab, both past and present, for their collegiality and friendship, including, but not limited to, Carla Anderson, Dr. Bjoern Gaertner, Brianne Patton, Elizabeth Canfield, and undergraduate students. I also acknowledge all scientists and staff of the Stowers Core Facilities and labs, whose excellent work made this project possible.

Finally, I am grateful for the enjoyable discussions with Dr. Man Mohan, Dr. Run Shen, Longhua Guo, Dr. Joey Nien, and whomever else I might be missing here, which kept me pursuing my dream when I got lost at midnights in the Stowers Institute.


embryo
genomic
pattern
DNA
pre-MBT
conserved
TSS
TBP
ZGA
enrichment
H3K9/14ac
expression
occupancy
Drosophila
motifs
activation
complex
established
studies
nuclei
roles
genome
gene
core promoter
binding
start
histone
ChIP-Seq
element
H3K4me3
results
nucleosome
transcriptional
proteins
chromatin
maternal
paused
associated
PcG
promoters
de novo
recruitment
embryogenesis
signal
cell
Pol II
development
activated
transcripts
active
zygotic
pausing
consistent
H3K27me3
factors
MBT
DI
transcription

Summary

During animal development, the first cell divisions of a fertilized egg are under maternal control. Zygotic transcription largely begins during the midblastula transition (MBT), but some genes may be transcribed earlier. How gene activation is accomplished is poorly understood. For example, it is unclear whether any pre-patterned markers, *e.g.* paused Pol II or histone modifications, are present at genes prior to activation.

In this study, I systemically investigated the dynamics of Pol II recruitment, histone modifications, and nucleosome accessibility on tightly staged *Drosophila* embryos to understand the establishment of zygotic transcription and chromatin organization during early embryogenesis.

Supported by evidence from histological assays, I found that the chromatin initially is loosely packed and there are no pre-recruited general transcription factors or pre-patterned histone modifications in the pre-MBT embryos for the global zygotic genome activation (ZGA). In addition, widespread Pol II pausing at developmental genes is established during the MBT for later activation while massive *de novo* Pol II recruitment occurs. Moreover, by comparing genes activated during MBT and ~110 genes strongly occupied by Pol II in the pre-MBT embryos, I found that the lack of Pol II pausing at pre-MBT genes correlates with strong core promoters that contain the TATA-box and binding motif of Zelda, a general activator recently identified for *Drosophila* ZGA.

Taken together, the function of core promoters might be an underappreciated mechanism for the general regulation of the *de novo* establishment of chromatin structure during early *Drosophila* embryogenesis. It is possible that this mechanism has evolved for adapting to the quick divisions in *Drosophila* early embryogenesis, and this may also be true for the ZGA of other vertebrate organisms with similar dividing patterns, *e.g.* zebrafish and *Xenopus*.

Abbreviations

AP	Anterior-posterior patterning
BRE	TFIIB recognition element
Cad	Caudal
CTD	carboxy-terminal domain
CDK	Cyclin-dependent kinase
ChIP	Chromatin immuno-precipitation
ChIP-Seq	ChIP coupled to high-throughput sequencing
Dl	Dorsal
DPE	Downstream promoter element
DRE	DNA replication-related element
DSIF	DRB-Sensitivity-Inducing Factor
DSP1	Dorsal switch protein 1
Dpp	Decapentaplegic
DV	Dorsal-ventral
ES cell	embryonic stem cell
GAF	GAGA factor
GRN	gene regulatory networks
GTF	general transcription factor
H3K4/K27me3	trimethylation of lysine 4/27 on histone H3
H3K9/14ac	acetylation of lysine 9/14 on histone 3
HAT	histone acetyltransferase
Hb	Hunchback
HMT	histone methyltransferase
Inr	Initiator element
MTE	Motif ten element
MZT	Maternal-to-Zygotic Transition
MBT	mid-blastula transition
NELF	Negative Elongation Factor
NT	nuclear transfer
PcG	Polycomb group
PIC	pre-initiation complex
PhoRC	Pho-repressive complex

PRC	polycomb repressive complex
PRE	PcG response element
P-TEFb	positive transcription elongation factor b
PHD	Plant Homeo Domain
Pol II	RNA polymerase II
RIPA	Radioimmunoprecipitation assay
SAGA	Spt-Ada-Gcn5-Acetyltransferase
SP1	Specificity factor 1
TBP	TATA box binding protein
TSS	transcriptional start site
TrxG	Trithorax group
TRE	TrxG response elements
TAF	TBP-associated factor
TFIID	transcription factor II D
ZGA	Zygotic Genome Activation

Table of contents

Table of Figures and Tables

CHAPTER 1

Introduction

Overview

Embryogenesis is a precisely regulated process that starts with a single fertilized egg. Despite the diversity of body patterns among various species, many early developmental events are highly conserved across different multi-cellular animals. One of those most intriguing events is the maternal-to-zygotic transition (MZT). Before conception, maternal proteins and RNAs are generously deposited into the oocyte during oogenesis. These maternally supplied apparatus can support initial development without turning on zygotic transcription. After that, maternally deposited RNAs are gradually degraded while Zygotic Genome Activation (ZGA) is triggered and begins to take over control from the maternal endowments. This process leads the totipotent cells of the early embryo to differentiate and execute their own specific roles during the upcoming gastrulation and organogenesis.

Within the past four decades, extensive studies have been performed in different organisms to understand how MTZ is regulated. Although many genes involved in this process have been identified, and several models have been proposed based on genetic and biochemical studies, it is still unclear how the chromatin changes and how it makes its contribution to transcription regulation. The limited number of nuclei in early embryos makes it extremely hard to investigate early chromatin dynamics. However, for some animals like *Xenopus*, zebrafish, and *Drosophila*, the global ZGA commences during mid-blastula transition (MBT), which occurs after 10-13 synchronous, gap-phase-free cell divisions and marks the transition to prolonged asynchronous division (Newport and Kirschner, 1982a; 1982b). This characteristic makes these animals ideal models for studying the chromatin state during the global ZGA with biochemical tools. Nonetheless, the way chromatin is organized during ZGA is still poorly understand. Specifically, this study is motivated by two questions. To begin with, has the early chromatin already been prepared for the ZGA, or is it programmed *de novo* during MBT? And, how is transcription achieved for the few genes that are expressed before MBT?

To address these questions, I first developed a system to sort pre-MBT and MBT embryos of *Drosophila melanogaster* for genomic study (Chapter 2). After collecting pure embryos in their desired stages, I systemically investigated the dynamics of Pol II recruitment, histone modifications, and nucleosome accessibility in the pre-MBT and MBT embryos with ChIP-Seq and nucleosome-mapping protocols that were modified to

accommodate limited material. Together with the histological results, the genomic data suggest that the chromatin status is gradually established during MBT. In addition, these high quality ChIP-Seq data for Pol II and TBP in pre-MBT embryos helped us identify ~110 genes that are activated before MBT. The over-representation of motifs associated with focused promoters in MBT genes suggests that the core promoter may play an important role in regulating the early transcription timing during *Drosophila* embryogenesis (Chapter 3). This study raises further questions about whether the *de novo* establishment of the early chromatin is in widespread existence during ZGA in different species, and also, what mechanisms contribute to this differential activation during early embryogenesis (Chapter 4). In this chapter, I will begin with an introduction of gene regulation in eukaryotic cells and how it can be affected by chromatin organization. This will be followed by an overview of early embryogenesis and the maternal-to-zygotic transition in *Drosophila*, and finally, a discussion on the aim of this study.

1.1 Gene regulation in eukaryotic cells

In a eukaryotic cell, the genomic DNA is packed by histones and non-histone proteins into chromatin, which was first named by Walther Flemming in 1880 because of its dyeability with aniline dyes (Olins and Olins, 2003). Chromatin is not only necessary for packaging the meter-length DNA into a ~10 μm-diameter nucleus, but it is also critical to the differential transcription of thousands genes in response to developmental or other environmental cues, especially for multi-cellular organisms.

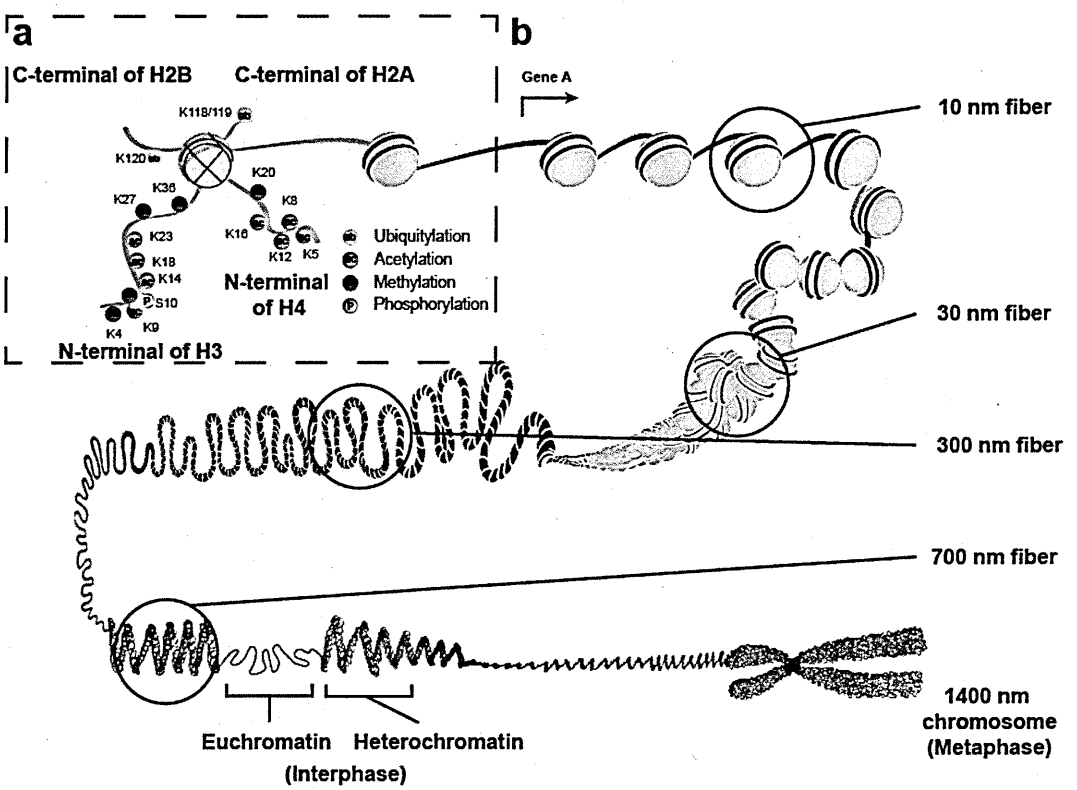


Figure 1-1. The hierarchical organization of chromatin.
a, Diagram of the nucleosome core particle with possible modifications. Genomic DNA is wrapped around an octamer formed by the four core histones H2A, H2B, H3 and H4, which have several amino acids that can be covalently modified. **b,** Compaction of DNA through packaging into nucleosomes. Incorporation of nucleosomes yields an 11 nm chromatin fiber ("beads on a string"). This structure is folded into a more compact fiber 30 nm in diameter, which can further fold into 300 nm and 700 nm fibers, and these can compact into chromosomes during mitosis. (Modified from (Pierce, 2009).

The fundamental structure of chromatin is the 10nm-diameter nucleosome, which is composed of a core histone octamer encircled ~1.65 times by ~146bp DNA (Kornberg, 1974). The octamer is composed of two molecules each of four highly conserved histone proteins: H2A, H2B, H3, and H4. Each core histone has an N-terminal tail, which can be covalently modified at different amino-acid sites (Figure 1-1a). These post-translation modifications, including acetylation, methylation, ubiquitination, sumoylation,

phosphorylation, etc., can affect chromatin configuration by either directly modulating DNA-nucleosome interactions or by changing the binding affinity of the nucleosome with other chromatin proteins (Kouzarides, 2007; Li et al., 2007).

With the linker histone, H1, and other non-histone proteins involvement, the 10nm-diameter fiber can form a 30nm diameter fiber (Figure 1-1b), which can be further organized into individual, condensed chromosomes during mitosis (Felsenfeld and Groudine, 2003). Chromatin is often organized into domains that are determined by the accessibility of DNA to transcription factors during interphase. These chromatin domains can be generally divided into two types: euchromatin and heterochromatin. The euchromatin is decondensed in interphase, highly accessible to transcription factors, and composed of actively transcribed genes. This permissive structure allows the euchromatin to replicate early in S-phase and remain highly sensitive to DNase digestion. In contrast, heterochromatin remains in a compact conformation during mitosis and is relatively inaccessible to transcription factors during interphase. Heterochromatin replicates late in S-phase and is resistant to DNase digestion (Grewal and Jia, 2007; Beisel and Paro, 2011). Although euchromatin and heterochromatin can usually be stably inherited between cell cycles (Probst et al., 2009), the regions containing regulated genes can be switched from one type of chromatin to another in response to developmental signals or other stimuli.

1.1.1 The assembly of general transcriptional machinery at the core promoter

Core promoter elements

The core promoter is a DNA sequence that frequently has low binding affinity to nucleosome, it directly interacts with the general transcriptional machinery, and it is located within ~100bp of the transcriptional start site (TSS). Based on the transcription initiation patterns, there are two major types of core promoters for the protein coding genes in metazoans. One is the focused promoter, where transcription initiates within a narrow region of a few nucleotides, and another is the dispersed promoter, composed of multiple weak start sites over a broad region of 50 to 100 nucleotides (Juven-Gershon and Kadonaga, 2010; Ni et al., 2010).

Focused promoters are frequently found within regulated genes. In the last forty years, intensive research on focused promoters has led to the identification of several sequence motifs such as TATA-box, Inr (Initiator element), DPE (Downstream promoter element), BRE (TFIIB recognition element), and MTE (Motif ten element). Also, research

has identified factors that bind specific core promoter elements (Ohler and Wassarman, 2010).

Although dispersed promoters are commonly used for housekeeping genes in animals and are prevalent in vertebrate genomes (around 70% of genes have dispersed promoters), the sequence motifs and their binding factors are still poorly understood. Among motifs found at dispersed promoters in *Drosophila*, the DRE (DNA replication-related element) is known to be bound by Dref (DNA replication factor), but factors that bind to other motifs [*e.g.* Ohler 1, 6, and 7 motifs (Ohler et al., 2002)] associated with dispersed promoters have yet to be identified. In addition, many genes have an intermediate initiation pattern, in which the transcription is broadly initiated but only a narrow region is preferentially used.

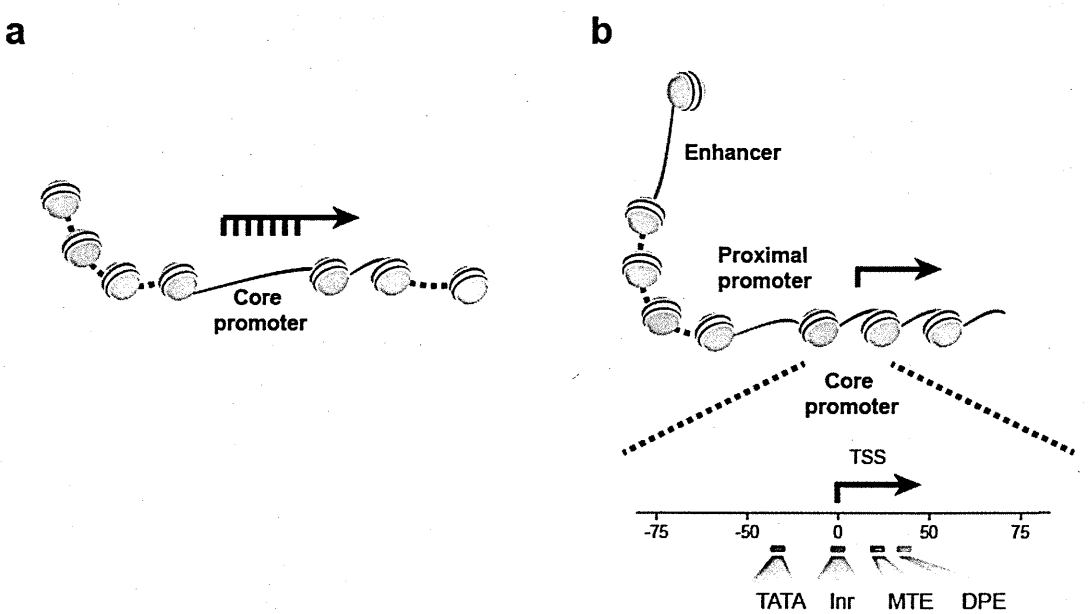


Figure 1-2. Two types of core promoters in eukaryotic cells.
a, Dispersed core promoters are typically found in housekeeping genes and usually yield multiple weak start sites over a region of 50–100 nucleotides. **b,** Focused core promoters are commonly used in regulated genes, in which transcription initiates within a few base pairs. Several core promoter motifs frequently occur in focused promoters.

RNA polymerase II and general transcription factors (GTFs)

There are three types of RNA polymerases, Pol I, Pol II, and Pol III, in charge of different sets of genes in eukaryotic cells. Pol I and Pol III transcribe ribosomal RNA, transfer RNA, and some small non-coding RNAs with little tissue bias (White, 2008). Pol

II transcribes all protein-coding genes and many non-coding RNAs (such as miRNA, etc.) (Lee et al., 2004), including regulated genes (Smale and Kadonaga, 2003).

Pol II has 12 highly conserved subunits that form a catalytic core with an unstructured extension from its largest subunit, RPB1. The unstructured extension, called the carboxy-terminal repeat domain (CTD), contains a tandem repeat of a conserved heptapeptide: $YS^2PTS^5PS^7$ (Corden, 1990). The repeat number varies between species, from 26 times in yeast, to 42 times in *Drosophila*, all the way to 52 times in human. These three serine residues are essential for cell viability (West and Corden, 1995) and can be phosphorylated during either transcription initiation by CDK7-cyclin H (S5 and S7) or during elongation by CDK9-cyclin T (S2). As such, modifications to S2, S5, and S7 serve as markers for studying Pol II dynamics in transcription regulation (Buratowski, 2009; Svejstrup, 2012).

Although Pol II is the functional unit for transcribing protein-coding genes, it does not have a DNA binding domain. Instead, it requires other general transcription factors (GTFs) to assemble a pre-initiation complex (PIC) at the promoter region to fulfill its function (Lee and Young, 2000). First, the TFIID complex needs to be assembled at the core promoter region through its DNA binding subunits. As the first identified DNA binding subunit of TFIID, the TATA box binding protein (TBP) can bind directly to the TATA box in a classic promoter, and then recruit TFIIB. TFIIB also has a DNA binding domain to recognize its core promoter motif, and it stabilizes TFIID-DNA binding. Once the TFIID/TFIIB complex is assembled at the promoter, it serves as a docking adaptor. Pol II, together with TFIIF and TFIIE, may then be recruited to the transcriptional start site either in a stepwise manner or pre-assembled as a holoenzyme. Then TFIIH, through its ATP-dependent helicase activity, unwinds 11–15 bp of DNA around the transcription start site to allow Pol II to initiate and transcribe. Meanwhile, the CDK7 subunit of TFIIH phosphorylates S5 within the heptad repeat of the Pol II CTD, which serves as a recognition site for recruiting mRNA capping machinery and histone methyltransferases, which mediate H3K4me3 (Thomas and Chiang, 2006).

1.1.2 The regulation of transcription initiation

Enhancers

In the absence of an activation signal, the core promoter of a regulated gene—despite its nucleosome-resistant property—might normally be occupied by nucleosomes

and fail to interact with GTFs (Zhang et al., 2009b; Bell et al., 2011). In this case, ,an enhancer is required for recruiting activators to the core promoter that make it accessible to GTFs in a signal dependent way (Lenhard et al., 2012). It can be a cis-regulatory element near the core promoter (*i.e.*, within 100-300 bp) or several kilobase pairs away from the core promoter (either upstream or downstream), and it can come into contact with the promoter via looping (Kadauke and Blobel, 2009; Ong and Corces, 2011). For some genes, even enhancers are wrapped into nucleosomes and are inaccessible to activators. In this case, a pioneer factor can bind to its targets, independent of nucleosome occupancy. This can change the conformation of local chromatin to make the enhancer accessible (Magnani et al., 2011; Zaret and Carroll, 2011).

Activators and coactivators

Most activators are small modular proteins containing both a DNA-binding domain, which recognizes enhancers, and activation domain, which mediates protein-protein interactions (Lee and Young, 2000). When triggered, activators bind to enhancers in a sequence-specific manner. They also recruit coactivators—which do not have DNA binding ability—to the core promoter to facilitate the assembly of the PIC.

There are three types of coactivators commonly required for transcription initiation (Näär et al., 2001): Mediator, chromatin modification complexes, and remodeling complexes. The Mediator is a highly conserved multi-subunit complex first identified as an adaptor for bridging activators and GTFs. In addition, it acts as a coactivator for looping enhancers together with the promoter, forming a chromatin hub to stabilize the PIC onto the core promoter (Malik and Roeder, 2010).

Most chromatin modification complexes recruited by activators serve as coactivators indirectly through histone modifications, although some of them can directly interact with GTFs. For example, the SAGA (Spt-Ada-Gcn5-Acetyltransferase) complex, a histone acetyltransferase (HAT), mediates histone 3 K9 and K14 acetylations (H3K9/14ac), which serve as dock sites for GTFs (like TAF1) or other coactivators containing bromodomain, including chromatin remodeling complexes (Weake and Workman, 2011). These chromatin remodeling complexes (*e.g.* the SWI2/SNF2 complex) use ATP as an energy source to slide or evict nucleosomes from the core promoter. In addition, H3K4me3, a histone modification that was previously found after TFIID assembly, has been recently reported to be directly recognized by TAF3 and to recruit

TFIID to the core promoter in mammalian cells (Vermeulen et al., 2007; Liu et al., 2011). This raises the possibility that some histone methyltransferases (HMTs) could also function as coactivators.

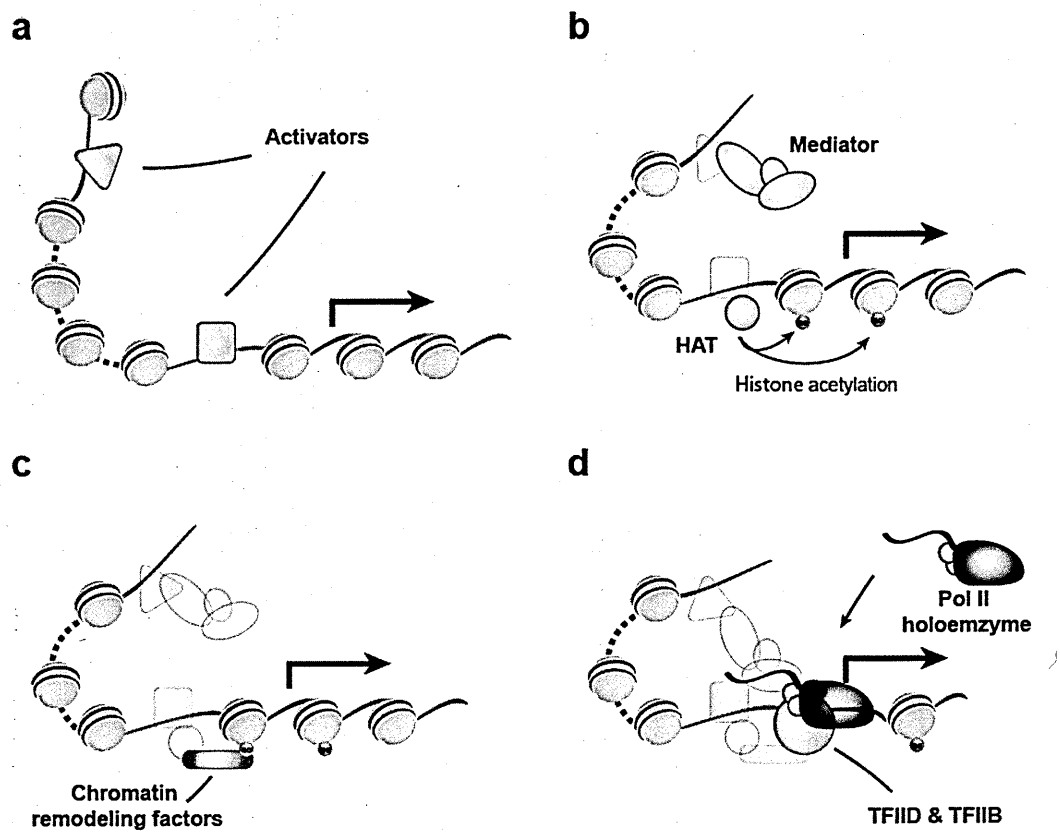


Figure 1-3. Regulation of Pol II recruitment at a regulated gene.
a, Activators bind to regulatory elements within enhancers. **b,** Through interactions with chromatin-modifying complexes, **c,** the promoter is then made accessible to Pol II by chromatin remodeling factors. **d,** Together with the mediator, general transcription factors are assembled at the core promoter and recruit Pol II holoemzyme.

1.1.3 Promoter proximal Pol II pausing

Although initiation has been proposed to be the key step for regulating gene activation, an accumulating body of research indicates that transcription elongation is also highly regulated. The initial studies of *Drosophila* heat shock genes (*Hsp* genes like *hsp70*) carried out in the Lis lab (Gilmour and Lis, 1985; 1986), found that even when GTFs are already assembled and transcriptionally engaged, they are unable to elongate in the absence of signal activation. Subsequent studies revealed that Ser 5 phosphorylated Pol II is paused at 20-60 bp downstream of the TSS of *Hsp* promoters with short nascent RNA (Rougvie and Lis, 1988; Rasmussen and Lis, 1993), and pausing often occurs with GAGA factor (GAF) bound to the *Hsp* promoter (Wilkins and Lis, 1997). This so-called promoter-proximal Pol II pausing (Pol II pausing) was later found in several other *Drosophila* genes,

mammalian oncogenes, and HIV genes (Adelman and Lis, 2012). Pol II pausing at un-induced genes suggests that it may play role in rapid transcriptional response to an activation signal.

Regulation of Pol II pausing and releasing

Two complexes are critical for establishing Pol II pausing both *in vitro* and *in vivo*: DRB-Sensitivity-Inducing Factor (DSIF) and the Negative Elongation Factor (NELF) complex (Wada et al., 1998a; Yamaguchi et al., 1999). DSIF is a heterodimer composed of Spt4 and Spt5, and it is highly conserved in all eukaryotes and archaea. DSIF can bind to the initiated transcription complex, which is an early step for blocking Pol II elongation and is required for subsequent recruitment of the NELF complex. The NELF complex is a four-subunit complex conserved from fly to human that is necessary for normal embryogenesis (Narita et al., 2003; Wang et al., 2010). A recent study suggests that knockdown of NELF-B in mouse embryonic stem (ES) cells can cause the premature derepression of developmental genes, leading the ES cells to spontaneous differentiation (Amleh et al., 2009).

Although DSIF and NELF together are sufficient to inhibit Pol II elongation *in vitro,* additional factors, like Gdown1—a Pol II binding protein, have been recently shown to be important factors for regulating Pol II pausing (Cheng et al., 2012). In addition, the combinatorial effects of tissue-specific transcriptional regulators may play a role in blocking Pol II elongation. Examples of these effects include the expression of gap genes in defining pair-rule genes' patterns (Wang et al., 2007), and also snail and twist repressing neural genes in *Drosophila* mesodermal tissue (Bothma et al., 2011). Moreover, nucleosome positioning, secondary structures of nascent RNA, and the core promoter sequence can also contribute to Pol II pausing (Narita et al., 2003; Hendrix et al., 2008; Mavrich et al., 2008; Eddy et al., 2011).

The activation of a gene with paused Pol II requires a signal-specific activator (*e.g.*, Myc or nuclear factor-κB), together with a bromodomain protein, BRD4, which further recruits the Positive Transcription Elongation Factor b (P-TEFb). P-TEFb is composed of one cyclin (cyclin T in *Drosophila*) and one cyclin-dependent kinase subunit, CDK9,

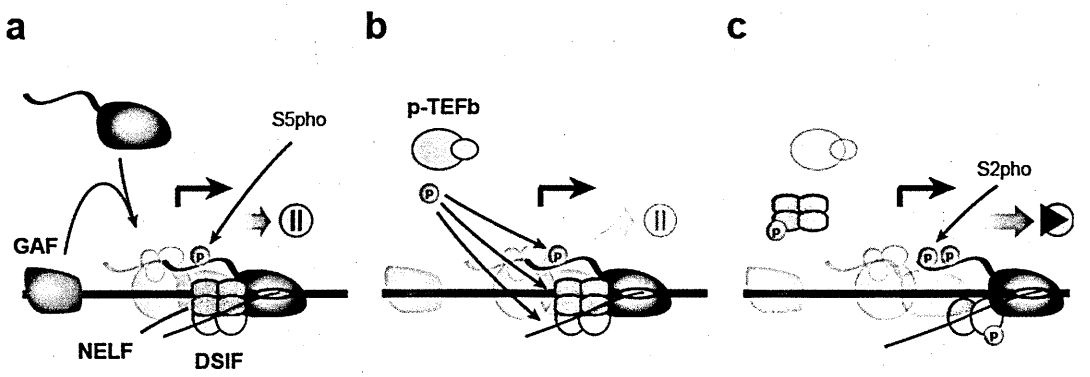


Figure 1-4. Regulation of Pol II pausing.
a, With the involvement of GAF, Pol II is recruited, but it remains paused by the action of NELF and DSIF. **b,** Once recruited to the promoter, pTEFb phosphorylates NELF, DSIF, and serine 2 of the Pol II CTD, **c,** which triggers the dissociation of NELF and the release of Pol II from the pause site. DSIF remains associated with Pol II and stimulates transcription elongation.

which can phosphorylate and release DSIF and NELF from paused Pol II (Wada et al., 1998b). P-TEFb can also phosphorylate serine 2 residues within the Pol II CTD repeats, which function as a platform for binding factors that facilitate RNA synthesis. Many of the coactivators mentioned above are also required for successful elongation. In addition, many HMTs have been associated with active genes. For instance, Set1 and Set2 are responsible for the H3K4me3 and H3K36me3 downstream of the core promoter and gene body, respectively (Adelman and Lis, 2012).

Widespread Pol II pausing in metazoans

Recent genomic studies in *Drosophila* and mammalian ES cells have revealed that Pol II pausing is widespread among developmental genes and stimulus-responsive genes (Guenther et al., 2007; Muse et al., 2007; Zeitlinger et al., 2007; Min et al., 2011). In addition to facilitating rapid responses to activation signals, recent evidence suggests that the paused Pol II may also be important for synchronizing the expression of some developmental genes (Boettiger and Levine, 2009) and for RNA processing (Buratowski, 2009).

Moreover, the widespread occurrence of Pol II pausing at quiescent developmental genes during embryogenesis indicates that the paused Pol II might reflect current transcriptional potential of these cells. However, it is unclear when widespread Pol II pausing is established during development.

1.2 Roles of chromatin in development

The developmental potential of embryonic cells irreversibly declines during animal embryogenesis. Researchers have been aware of this transition since the late 1900s, when Hans Spemann performed his serial experiments on salamander early embryos to find an answer for the long-standing Preformation-Epigenesis debate. The fading competence of embryonic cells during development was later described by Conrad Hal Waddington as an epigenetic landscape, in which differentiating cells behave like marbles rolling into the lowest points on a surface while the slopes become steeper as development progresses (Waddington, 1957).

It was later shown, for most animals tested, that this naturally irreversible process is not due to permanent alterations of the genomic DNA because viable animals can be obtained from differentiated cells using somatic cell nuclear transfer (NT) experiments (Gurdon, 1962; Wilmut et al., 1997; Wakayama et al., 1998). However, the NT efficiency decreases as donor cells differentiate, which suggests a level of regulation beyond genetics called epigenetics (Gurdon and Melton, 2008). Epigenetic control is caused by heritable changes in gene function that cannot be explained by changes in DNA sequence (Holliday, 1994). Chromatin organization could be primarily responsible for this control.

Chromatin dynamics, including histone modifications, are under the control of tissue-specific regulators but can be maintained during development after these tissue-specific regulators are no longer present (Goldberg et al., 2007; Kouzarides, 2007). In most cases, histone modifications recruit non-histone proteins containing specific domains, such as a bromodomain for acetylations, and chromo, tudor, MBT, and PHD domains for methylation. The recruitment of these non-histone proteins can initiate the assembly of large protein complexes. These may contain co-factors involved in gene regulation and enzymes that maintain the modifications during cell division. Two conserved groups of chromatin proteins, Polycomb group (PcG) proteins and Trithorax group (TrxG) proteins, represent the most well studied epigenetic regulators of metazoan development.

1.2.1 PcG proteins mediated gene silencing

The formation of heterochromatin is an epigenetic change that can be stably inherited during division. Heterochromatin can be subdivided into constitutive and facultative heterochromatin. Constitutive heterochromatin is usually composed of highly repetitive DNA with few genes and is mostly located in centric and telomeric regions of

chromosomes, marked by H3K9me3. It exists in all tissue types to repress transposable element activity and to maintain the genomic integrity. Facultative heterochromatin does not have as much repetitive DNA and is more tissue-specific. It can de-condense for transcription under specific developmental or environmental signaling cues.

One of the most well-studied mechanisms that establishes facultative heterochromatin is Polycomb-mediated silencing, which was first uncovered in Drosophila (Lewis, 1978; Jürgens, 1985) and is highly conserved from *C. elegans* to mammals. The function of Polycomb group (PcG) proteins is to maintain the repressed state of Hox genes and many other developmental genes to keep the cell-lineage identity stable and heritable.

In *Drosophila*, biochemical studies suggest that PcG proteins can form three major complexes, polycomb repressive complexes 1 and 2 (PRC1 and PRC2), and Pho-repressive complex (PhoRC) (Schwartz and Pirrotta, 2007; Simon and Kingston, 2009). Despite decades of research on PcG proteins, their mode of action is still not entirely understood. I will explain one widely described model.

The first step in the assembly of the PcG complexes is the recruitment of PhoRC to a DNA region called PcG response element (PRE). A typical PRE has a binding site for Pleiohomeotic (Pho) or Pleiohomeotic-like (Phol), which are the only subunits of PhoRC that have the DNA binding domain. In addition, the PRE also has several motifs recognized by various DNA binding proteins such as GAF, Zeste, Dorsal switch protein 1 (DSP1), Pipsqueak, Grainyhead, and Specificity factor 1 (SP1), which might function as co-recruiters for PRC1 or PRC2.

The binding of PhoRC further recruits PRC2. Together with two other components (Esc and Su(z)12) of the PRC2, Enhancer of *zeste* [E(z)] tri-methylates H3K27 on the surrounding nucleosomes directly and may spread H3K27me3 to more distant sites by looping. H3K27me3 can then serve as a docking site for PRC1 assembly. Polycomb (Pc), one component of the PRC1, binds to the H3K27me3 through its chromo domain and recruits other components of the PRC1, which can mediate H2A mono-ubiquitination (K118 in *Drosophila* or K119 in mammals), and finally carry out its repression function, presumably by chromatin compaction (Francis et al., 2004).

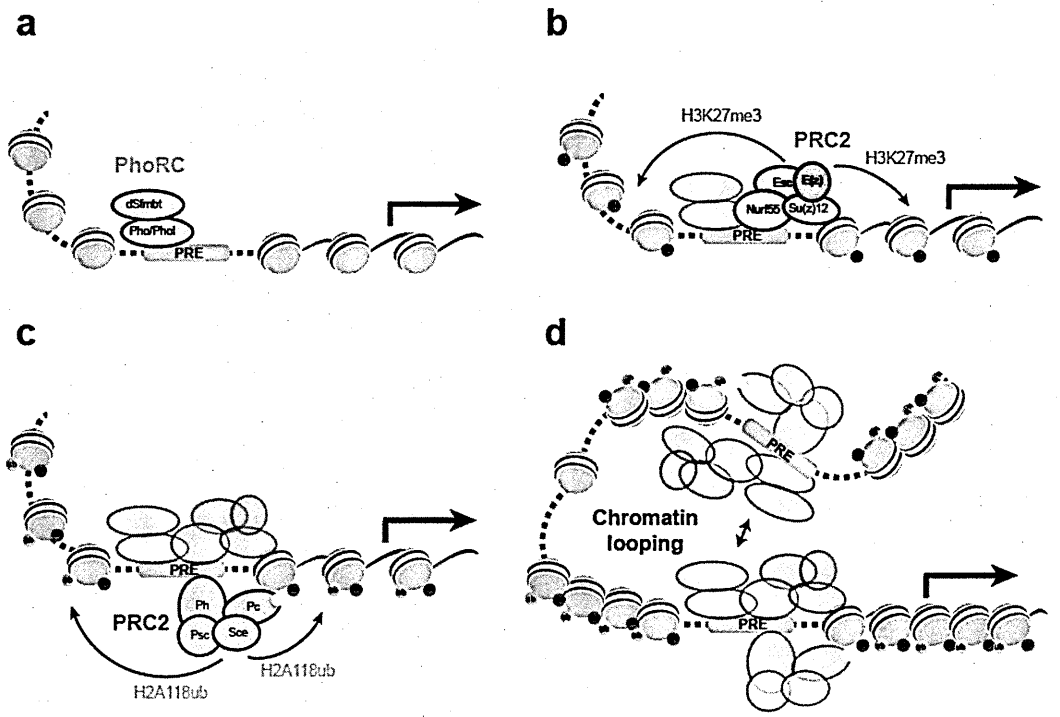


Figure 1-5. A stepwise model for PcG mediated silencing.
a, PhoRC binds to PREs (through Pho or Phol), and recruits PRC2, which catalyzes H3K27 trimethylation on flanking nucleosomes (**b**). **c,** PRC1 is recruited around PREs due to its Pc subunit's affinity for H3K27me3, and it mediates H2A ubiquitination at nearby nucleosomes. **d,** Binding of PRE-localized Pc to K27me3 might subsequently lead to chromatin looping and compacting.

1.2.2 TrxG proteins, antagonists of PcG proteins

In contrast to the silencing roles of PcG proteins, TrxG proteins were first found to be critical for maintaining the active status of Hox genes in *Drosophila* embryogenesis when early activators disappear in later developmental stages (Ingham, 1983). It was later shown that TrxG proteins function as antagonists of PcG-mediated silencing and are involved in ES cell self-renewal, cell fate choice, proliferation, and tumorigenesis by maintaining active chromatin states.

TrxG proteins can be divided into three classes based on their molecular functions. Proteins of the first group are directly recruited to TrxG response elements (TREs). Factors required for PcG proteins recruitment, like pho and GAF, are also critical for the Trx complex assembly, which cause TREs to coincide with PREs in most cases. Another group of proteins, SET domain-containing factors can further be recruited to form COMPASS or a COMPASS-like complex and deposit H3K4me3 to their target chromatin. This may occur with the assistance of non-coding RNA (Sanchez-Elsner et al., 2006). UTX, a histone demethylase (specifically demethylates H3K27me3), which has been recently identified as one subunit of the COMPASS-like complex in many model organisms (Lan et

al., 2007; Lee et al., 2007; Smith et al., 2008; 2011) and may play an important role in releasing PcG-mediated silencing. Through the H3K4me3, ATP-dependent chromatin remodeling factors, such as ASH1 and BRM, can be recruited to regulate the transcription activity of their targets. Unlike other TrxG proteins, which can be recruited to the target genes independently of their activation status (Papp and Müller, 2006), ASH1 and BRM are targeted to genes in an activation-dependent manner (Déjardin and Cavalli, 2004).

1.2.3 Bivalent domain, a chromatin marker for gene plasticity in development

Despite the counteracting roles of PcG and TrxG proteins during development, recent genomic studies in mammalian ES cells and adult stem cells found that promoters of many quiescent developmental genes are co-occupied with the marks of these two protein groups, specifically H3K27me3 and H3K4me3 (Bernstein et al., 2006; Mikkelsen et al., 2007; Pan et al., 2007; Zhao et al., 2007; Cui et al., 2009; Lien et al., 2011). These so-called "bivalent domains" are released during differentiation: they either lose H3K27me3 at genes activated by appropriate signals, or they lose H3K4m3 at genes that need to be silenced during fate determination. This bivalent state is similar to the widespread paused Pol II at the TSS of developmental genes in ES cells and may reflect the transcriptional plasticity of those developmental genes (Guenther et al., 2007; Min et al., 2011). Interestingly, so far there is no bivalent domain reported in *Xenopus* and *Drosophila*, although they both contain paused Pol II in early embryogenesis (Akkers et al., 2009; Schuettengruber et al., 2009).

One intriguing question is when H3K4me3 and H3K27me3 markers are established during development. Studies in mammals and zebrafish have shown that there are bivalent domains at developmental genes in sperm (Hammoud et al., 2009; Brykczynska et al., 2010; Wu et al., 2011), and this raises the possibility that those markers can be inherited from gametes. However, there is neither functional evidence to support this hypothesis, nor is there any link from the gametes to ES cells or zebrafish embryos (Vastenhouw et al., 2010). Therefore, it is difficult to conclude whether those markers are inherited or established *de novo* during embryogenesis. In addition, the connection between bivalent domains and Pol II pausing is still an open question.

1.3 Early embryogenesis of *Drosophila melanogaster*

The fruit fly, *Drosophila melanogaster,* has been adapted by geneticists as a model organism for more than a century (Rubin, 2000), not only because of its simple husbandry, but also its small genome and remarkable exoskeleton features are ideal for phenotype screening. Although it is an invertebrate with a short life cycle, many developmental processes are conserved between *Drosophila* and vertebrates [*e.g.*, the Ca^{2+} signaling in oocyte activation (Horner and Wolfner, 2008), the BMP pathway in dorsal-ventral patterning (Martindale, 2005), and *Hox* clusters in anterior-posterior patterning (Garcia-Fernàndez, 2005)]. Moreover, many conserved developmental genes only have one homologue in *Drosophila*, which makes studying developmental events and gene functions much easier than other more complicated vertebrate systems. Together with current transgenic tools, studies of mutants with developmental defects obtained over decades of screenings have built up our current understanding of *Drosophila* embryonic development.

1.3.1 Overview of Early *Drosophila* embryogenesis

In *Drosophila*, the mature MI-phase arrested oocyte is activated during ovulation, a few minutes before fertilization. Mechanical stimulation during ovulation, rather than fertilization, is critical for initiating the resumption of meiosis (Doane, 1960), protein synthesis (Driever et al., 1988), and mitotic oscillation (Theurkauf and Hawley, 1992). However, the fact that unfertilized eggs from most *Drosophila* species, especially *melanogaster*, are hardly developed—even through parthenogenesis—suggests that fertilization is required for early development (Heifetz et al., 2001).

Once two pronuclei fuse after fertilization, the one-cell embryo starts its unique cleavage stage (also called the syncytial blastoderm stage) during which 13 cycles of synchronous nuclear divisions occur without forming cells (Sullivan and Theurkauf, 1995). The first 8 divisions are gap phase free and only take one hour in total. After that, 256 nuclei are produced and start to migrate to the periphery of the embryo. Meanwhile, the 5 nuclei to reach the posterior surface first form the pole cells with cell membranes, which differentiate into germ cells later. After another hour, 4 rounds of division with short G2 phases are completed, the division dramatically slows down, and the embryo starts cellularization by folding the oocyte plasmid membrane inward and fusing intra-embryonic vesicles to form the cell membranes for all peripheral nuclei (Loncar and Singer, 1995;

Sokac and Wieschaus, 2008). This causes the embryo to transition from a syncytial blastoderm to a cellular blastoderm. Shortly after cellularization, the embryo divides asynchronously and starts gastrulation to form the three germ layers (ectoderm, mesoderm, and endoderm) with anterior-posterior identity. These later differentiate into the different cell types that make up the individual.

1.3.2 Gene regulatory networks for body patterning

Although cellular identity in *Drosophila* embryos is determined during cellularization, the gene regulatory networks (GRNs) for the body patterning are initiated by maternally-loaded morphogens during oogenesis and are established during the syncytial blastoderm stage.

Dorsal-ventral patterning

Dorsal-ventral (DV) patterning is predominantly controlled by the activity of Dorsal (Dl, the *Drosophila* homolog of mammalian NF-kB). This activity is defined by the asymmetric activation of the EGF pathway on the dorsal follicle cells during oogenesis (Morisato and Anderson, 1995). In the ventral follicle cells, the restricted expression of Pipe—a sulfortransferase which is repressed by EGF signal—triggers a serine protease cascade in the ventral region of the perivitelline space. Finally, it activates the Toll receptor of the embryonic membrane through its ligand Spätzle (Spz). This leads to the degradation of Cactus and the transportation of Dl into the embryonic nuclei in a gradient going from ventral to dorsal.

The nuclear Dl gradient sets the dorsal-ventral axis through three basic transcriptional responses (Hong et al., 2008). [1] In the most ventral part of the embryos, high levels of nuclear Dl bind to low-affinity targets and activate genes like *twist* and *snail*. [2] At the ventral lateral sides, intermediate levels of Dl along with low levels of *twist* activate high affinity Dl target genes (*e.g.*, *vnd* and *brk*). [3] At the dorsal lateral sides, low levels of Dl bind to their high-affinity targets, leading to the activation of genes for neuro-ectoderm formation (*e.g.*, *sog* and *ths*).

Meanwhile, at the dorsal side of the embryo without nuclear Dl activity, Decapentaplegic (Dpp) is expressed and activates the *Drosophila* BMP signaling pathway (St Johnston and Gelbart, 1987). This results in the activation of some dorsal ectodermal genes (like *zen* and *tld*) and the repression of other Dl targets, like *brk*. Moreover, cross-

talk between the Dl and other signaling pathways (Hong et al., 2008), such as Notch and EGF, can further regulate other downstream genes. This GRN determined by the Dl activity gradient finally subdivides the embryo into amnioserosa, epidermis, neurogenic ectoderm, and mesoderm along the DV axis at the end of cellularization.

Anterio-Posterior patterning

During oogenesis, *bicoid (bcd)* and *nanos (nos)* are maternally loaded as mRNAs and are docked to the anterior and posterior poles of the oocyte, respectively, by dynamic transportation along the polarized microtubules (Roth and Lynch, 2009). After fertilization, translation of *bcd* and *nos* generates two complementary morphogen gradients in the syncytium. Bcd diffuses from anterior to posterior, and Nos diffuses from posterior to anterior. Bcd and Nos further block the translation of two ubiquitously loaded mRNAs, *hunchback (hb)* and *caudal (cad)*. This leads the formation of two additional morphogen gradients in the syncytium: an anterior to posterior gradient of Hb and a posterior to anterior gradient of Cad (going from high to low concentration).

Gradients of three transcription factors, Bcd, Hb, and Cad, in the syncytial blastoderm initiate the Anterio-Posterior (AP) patterning by activating or repressing the gap genes, like *knirps* and *giant*, into broad regions along the AP axis. Products of gap genes are also transcription factors. Together with Bcd, Hb, and Cad, the combined regulatory action of the gap gene products activates the pair-rule genes in periodic striped patterns just before cellularization. The alternative expression of pair-rule genes in transverse strips leads to the formation of 14 parasegments and the activation of segmentation genes. Meanwhile, products of gap genes and pair-rule genes together regulate Hox genes, which determine the segment identity for larvae and adults along the AP axis at the end of cellularization (Rivera-Pomar and Jäckle, 1996; Peel et al., 2005).

1.3.3 Maternal-to-Zygotic Transition

In multicellular animals, maternally supplied apparatus (*e.g.* mRNAs and proteins) are generously loaded not only for supporting the fertilized egg's early divisions and metabolism without necessitating transcription of its own zygotic genes, but also for patterning its early development.

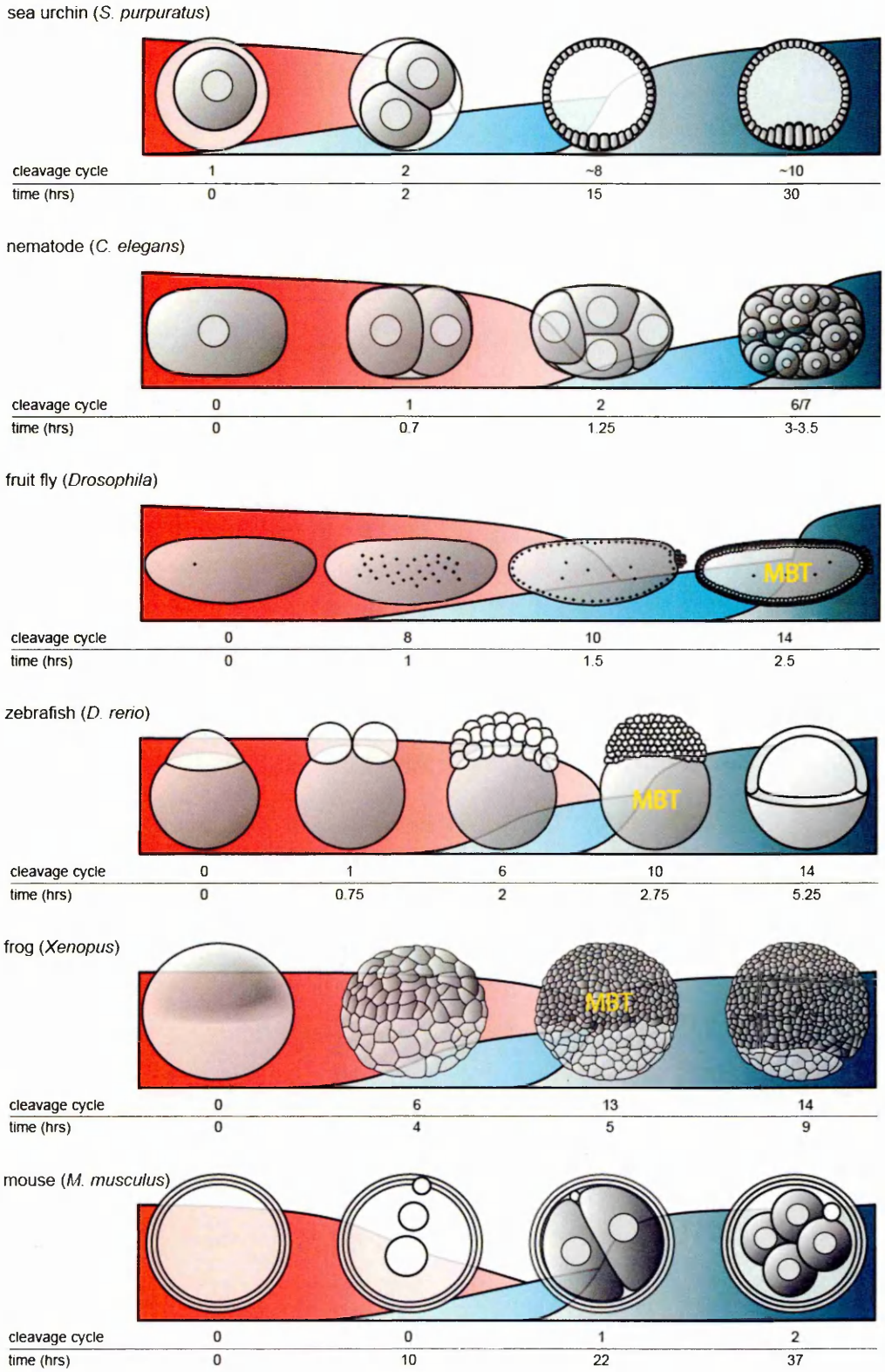


Figure 1-6. Maternal-to-Zygotic Transition is conserved across different model animals. Despite different cleavage patterns after fertilization, the key events of MZT are well-conserved: Along with the degradation of maternal RNA (shown in red), a few zygotic genes begin transcription (marked by light blue), and are followed by global zygotic genome activation (marked by dark blue). For *Drosophila*, zebrafish, and *Xenopus*, the global zygotic genome is activated during mid-blastula transition (MBT, marked in yellow). (This figure is modified with permission from Dr. Howard D. Lipshitz.) (Tadros and Lipshitz, 2009)

To further differentiate cells to perform distinct roles in development, the embryo needs to set up its own zygotic transcription and take over developmental control by actively clearing maternal transcripts. The timeline of the maternal-to-zygotic transition (MZT) varies among different organisms: from the first two mitotic divisions (sea urchin, mammals and *C. elegans*) to several gap-phase free divisions after fertilization (like *Drosophila*, zebrafish and *Xenopus*). However, the main events of MZT are generally similar and include the gradual degradation of maternally deposited RNAs after egg activation and zygotic genome activation (ZGA) (Tadros and Lipshitz, 2009).

Maternal RNA degradation

In *Drosophila*, although maternally supplied material can support embryonic development until the beginning of cellularization without zygotic transcription (Arking and Parente, 1980; Gutzeit, 1980; Edgar and Schubiger, 1986), maternal RNAs start to degrade at cycle 8, which coincides with the transcription of early zygotic genes. One RNA-binding protein, Smaug (SMG), translated by maternal *smaug* mRNA upon egg activation, is critical for degrading most maternal mRNA because it recruits the deadenylase complex to its targets (Tadros et al., 2007). Recent work has found that zygotic miRNA families, like the miR-309 cluster, and piRNA are involved in the degradation of several hundred maternal mRNAs in a SMG dependent manner (Bushati et al., 2008; Rouget et al., 2010).

Zygotic Genome Activation

In *Drosophila*, the bulk of zygotic genome activation occurs after division 13, when the embryo is starting cellularization (Lamb and Laird, 1976; Anderson and Lengyel, 1979). This window has also been called the mid-blastula transition (MBT), since it is similar to the MBT which was first described in *Xenopus* (Newport and Kirschner, 1982a; 1982b) and zebrafish (Kane and Kimmel, 1993), in terms of the massive zygotic genome activation that co-occurs with the transition from synchronous divisions to prolonged asynchronous divisions. In addition to most housekeeping genes [except the histone cluster, (Edgar and Schubiger, 1986)], developmental genes involved in pattern formation, sex determination, and gastrulation, are transcribed within this window. This is consistent with the fact that only a few zygotic mutants have been identified in the large-scale screens for genes that affect early development before cycle 14 (Jürgens et al., 1984; Nüsslein-Volhard et al., 1984; Wieschaus et al., 1984).

However, similar to *Xenopus* and Zebrafish (Kimelman et al., 1987; Nakakura et al., 1987), the mRNA of a handful of zygotic genes were visible in RNA *in situ* experiments before *Drosophila* MBT. Although some short nascent transcripts have been found in the embryos before cycle 10 by electron microscopy (McKnight and Miller, 1976), there is no evidence that any known genes are transcribed in the pre-blastoderm embryos. The first identified pre-MBT genes were found through a screen for genes that affect body patterning. These pre-MBT genes—including most gap genes (*h, hb,* and *gt*), some pair-rule genes (*ftz* and *eve*), and DV genes (*snail* and *zen*)—are directly activated by maternal morphogens (like bicoid and dorsal) and determine the expression patterns of genes activated during MBT (Bosch et al., 2006). Other pre-MBT genes have been identified as functional proteins involved in cellularization (like *nullo*, *bnk,* and *sry-a*) and sex-determination (such as *Sxl* and *sis-A*), which is consistent with their later roles in upcoming developmental events. Besides those developmental genes, histone clusters (an exception among housekeeping genes) are also actively transcribed in the pre-MBT embryos.

Timing the onset of ZGA

The step-wise nature of ZGA during *Drosophila* early divisions raises the possibility that the onset of the ZGA may be regulated by some titratable repressors in a nucleocytoplasmic ratio (N/C ratio) dependent manner (Brown et al., 1991). Indeed, the transcription of *ftz* is N/C ratio sensitive and is repressed by Tramtrack (TTK), a maternally deposited repressor (Pritchard and Schubiger, 1996). However, by systematic manipulation of DNA content, only a small portion of zygotic genes have been found to be influenced by the N/C ratio during early embryogenesis (Lu et al., 2009). The fact that most new zygotic transcripts correlate with absolute time suggests the existence of a maternal clock. This may be regulated by SMG mediated maternal mRNA degradation (Benoit et al., 2009).

In addition, the absence of transcription during the early embryogenesis may be due to the rapid cell cycles and the abortion of transcription during rapid DNA replication. This idea is supported by the application of cell cycle inhibitors in the embryos after nuclear division 10, which results in the premature ZGA (Edgar and Schubiger, 1986). This is also consistent with the finding that most early transcripts are short: miRNAs and small protein-coding genes with few introns (De Renzis et al., 2007). However, the failure to induce premature ZGA with cell cycle inhibitors before cycle 10 suggests that rapid nuclear divisions alone are not sufficient for preventing ZGA during early embryogenesis.

The last model for the timing of ZGA hypothesizes that the initial chromatin is not competent for transcriptional activation. It could be either lacking some general activator or coactivator, as is found in mouse embryogenesis (Bultman et al., 2006). But this model cannot explain why there are differential onsets of ZGA during embryogenesis, especially the activation of pre-MBT genes before the massive ZGA during MBT.

1.4 Aims of this thesis

Understanding chromatin organization during early *Drosophila* embryogenesis

Drosophila embryogenesis is an appropriate system for investigating transcription regulation on a genome-wide level. As mentioned, there is a large body of knowledge from extensive mutant screens and functional studies of interactions between trans-regulatory factors and cis-regulatory elements. Furthermore, the small and well-annotated genome and the availability of large quantities of *in vitro* developing embryos greatly facilitate genome-wide studies (Kharchenko et al., 2011; Nègre et al., 2011; Bonn et al., 2012). However, studying ZGA in *Drosophila* with genome-wide approaches is not entirely straightforward because *Drosophila* females occasionally hold their eggs longer than usual before deposition. This produces a small degree of contamination in the form of out-of-stage embryos. Since these older embryos have substantially more nuclei during the early stages, such contamination is significant. The numerous overlooked nuclei from out-stage embryos may make results hard to interpret.

In this study, my aim was to develop an experimental setup that overcomes the hurdle of later-stage contamination for genomic studies. With such a setup, the goal was to understand the dynamic chromatin organization during *Drosophila* early embryogenesis. Specifically, my questions were:

1) When is Pol II pausing established, and does it contribute to the widespread zygotic genome activation?

2) Are there any histone modifications that serve as pre-patterning markers for zygotic genome activation?

3) How accessible is the chromatin during early embryogenesis when transcription is minimal?

CHAPTER 2

Materials and Methods

2.1 Materials

2.1.1 Chemicals and Reagents

DAPI	Sigma (32670)
Agarose	Sigma (A5093-500G)
Biotin-14-dATP	Invitrogen (19524-016)
bleach	Clorox (A32235 IL-1)
β-mercaptoethanol	Sigma (M6250)
Bovine serum albumine (BSA)	Sigma (B9001S)
Chloroform	Sigma (C2432)
Protease Inhibitor Cocktail Tablets, EDTA-free	Roche (11873580001)
Dextran sulfate sodium salt	Sigma (42867)
Dithiothreitol (DTT)	Sigma (P117B)
dNTP Mix (10 mM)	Promega (U1511)
Ethanol	Sigma (2716)
EDTA	Sigma (AM9260G)
EGTA	Sigma (E3889)
Fast SYBR Green Master Mix	AB (4385612)
Formaldehyde (37% solution)	JT Baker (2106-04)
Formamide	Sigma (F7503-500ML)
GelRed nucleic acid gel stain	Biotium (41003)
Glycogen	Roche (10901393001)
Heparin sodium salt	Sigma (84020)
Heptane	Sigma (H2198-4L)
Igepal CA-630 (NP-40)	Sigma (I8896)
Isopropanol	Sigma (BP2632-4)
Lithium chloride (8M solution)	Sigma (L7026-500ML)
Magnesium chloride	Sigma (M8266)
Methanol	FisherScientific (A452SK4)
N-lauroylsarcosine sodium salt	VWR (TCL0019-025G)
Phenol	Sigma (P4682)
Phenol-chloroform-isoamyl alcohol (25:24:1)	Sigma (77617)
Potassium chloride	Sigma (P9541-500G)
Potassium dihydrogen phosphate	JT Baker (BDH3447-1)

Salmon sperm DNA	Invitrogen (15632-011)
Sodium acetate	Sigma (S2889)
Sodium azide	Fluka (S2002)
Sodium chloride	Sigma (S3014-1KG)
Sodium deoxycholate	Sigma (D6750-500G)
Sodium dodecyl sulfate (SDS, 10% solution)	Invitrogen (24730-020)
Sodium hydrogenphosphate	VWR (JT3822-7)
Spermidine	Sigma (S2626-5G)
Tris (hydroxymethyl) aminomethane (Tris)	Sigma (252859)
Trisodium citrate	Sigma (S1804)
Triton X-100	Sigma (T8787-250ML)
VECTASHIELD Mounting Medium	Vector lab, Inc (H-1000)

2.1.2 Buffers

Phosphate buffered saline (PBS)

137 mM NaCl; 2.7 mM KCl; 10 mM Na_2HPO_4; 2 mM KH_2PO_4

Autoclave and kept at RT

Tris-acetate-EDTA buffer (TAE, 50×)

242 g Tris; 100 mL EDTA (0.5 M stock, pH 8); 57.1 mL glacial acetic acid; added H_2O to 1L, filter sterilized and kept at RT

Saline-sodium citrate buffer (SSC, 20×)

3M NaCl; 300mM trisodium citrate (adjust pH to 7.0)

filter sterilized and kept at RT

Crosslinking Buffer

50 mM Hepes; 1 mM EDTA; 0.5 mM; EGTA; 100 mM NaCl

filter sterilized and kept at RT

Buffer A1

15 mM HEPES, pH 8; 15 mM NaCl; 60 mM KCl; 4 mM $MgCl_2$; 0.5% Triton X-100; 0.5 mM DTT

Filter sterilized and kept at 4°C

Buffer A2

15 mM HEPES, pH 8; 140 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 1% Triton X-100; 0.1% sodium deoxycholate; 1% SDS; 0.5% N-lauroylsarcosine

Filter sterilized and kept at 4°C

RIPA buffer

50 mM HEPES, pH 7.5; 1 mM EDTA; 0.7 % sodium deoxycholate; 1 % NP-40; 500 mM LiCl

Filter sterilized and kept at 4°C

TE

10 mM Tris-HCl, pH8; 1 mM EDTA

Filter sterilized and kept at RT

TE + salt

10 mM Tris, pH 8.0; 1 mM EDTA; 50 mM NaCl

Filter sterilized and kept at RT

Elution buffer

50 mM Tris, pH 8.0; 10 mM EDTA; 1 % SDS

Filter sterilized and kept at RT

NPS buffer

0.5 mM Spermidine; 0.075% IGEPAL; 50 mM NaCl; 10 mM Tris-Cl, pH 7.5; 5 mM $MgCl_2$; 1 mM $CaCl_2$; 1 mM β-mercaptoethanol

Filter sterilized and kept at 4°C

Binding & Washing buffer

5 mM Tris-HCl, pH 7.5; 0.5 mM EDTA; 1 M NaCl

Filter sterilized and kept at RT

NEB buffer 2

10 mM Tris-HCl, pH 7.5; 50 mM NaCl; 10 mM $MgCl_2$

Filter sterilized and kept at RT

P1 solution

50mM Tris, pH 8; 10mM EDTA; 100 ug/ml RNase A

Filter sterilized and kept at 4°C

P2 solution

200mM NaOH; 1% SDS

Filter sterilized and kept at RT

P3 solution

3M KOAc, pH 5.5

Dissolve 58.9g to 150ml ddH_2O, adjust to pH5.5 with glacial acetic acid and add ddH_2O to 200ml. Autoclave and store at 4°C

FISH Hybridization Buffer

2× SSC; 10% dextran sulfate; 50% formamide; Salmon sperm DNA at 0.5 mg/ml (sonicated 15min by Bioruptor)

It can be stored at -20°C for long time. Warm to 37°C before use

pre-Hybridization Mixture:

50% formamide; 4× SSC; 100 mM NaH_2PO_4, pH 7.0; 0.1% Tween 20

Fixative A

25% glutaraldehyde (EM grade); 50 mM sodium cacodylate (pH 7.0)

Fixative B

2% glutaraldehyde (EM grade); 50 mM sodium cacodylate (pH 7.0)

Fixative C

1% osmium tetroxide; 2% glutaraldehyde; 50 mM sodium cacodylate (pH 7.0)

2.1.3 Kits

FISH Tag DNA Multicolor Kit	Invitrogen (F32951)
MinElute PCR Purification Kit	QIAGEN (28006)
QIAquick Gel Extraction Kit	QIAGEN (28706)
QIAquick PCR Purification Kit	QIAGEN (28106)
Paired-End DNA Sample Preparation Kit	Illumina (PE-102-1001)
Terminal Transferase end labeling Kit	Roche (03333574001)
Promega PCR clean-up kit	Promega (A9282)

2.1.4 Affinity purification material

Dynabeads M-280 Sheep Anti-Mouse IgG	Invitrogen (11201D)
Dynabeads M-280 Sheep Anti-Rabbit IgG	Invitrogen (11203D)
Dynabeads MyOne Streptavidin C1	Invitrogen (65001)

2.1.5 Enzymes and antibodies

Micrococcal nuclease	Worthington Biochem (LS004798)
Proteinase K	Invitrogen (100005393)
RNase A	Sigma(R6513)
Rabbit anti-H3K4me3	Cell Signaling Technologies (9751S)
Rabbit anti-H3K27me3	Active Motif (39155)
Rabbit anti-H3K9/14ac	Millipore (07-360)
Mouse anti-RNA Pol II (8WG16)	Covance (MMS-126R)
Mouse anti-RNA Pol II (CTD4H8)	Millipore (05-623)
Rabbit anti-dTBP	Gift from Dr. James Kadonaga
Rabbit anti-Polycomb	Santa Cruz (sc-25762)
Mouse anti-lamin 0	DSHB (ADL67.10)
Alexa Fluor 488 chicken anti-rabbit IgG	Invitrogen (A-21441)
Alexa Fluor 488 goat anti-mouse IgM	Invitrogen (A-21042)
Alexa Fluor 555 donkey anti-mouse IgG	Invitrogen (A-31570)

2.1.6 Fly genomic clones

ANT-cluster (CH321-21L11, CH321-76C24)

BX-cluster (CH321-78J17)

(BACs were ordered from Children's Hospital Oakland Research Institute)

2.1.7 Fly stocks

Oregon R (wild-type) was obtained from the Bloomington *Drosophila* Stock Center at Indiana University

gd^7 was kindly provided by Michael Levine, University of California at Berkeley

2.1.8 Oligos ordered from IDT

NonG-f	(Negative control)	ATTGCTGCATCTTTGGGATG
NonG-r		TCGTGAAATGTTTGCTACTGGA
Sna-f	(for pre-MBT embryos)	AAATGTCAATTTGAGCAATGG
Sna-r		ATCTGCTCGCACGCACTTAC
Opa-f	(for MBT embryos)	GCCACGTACCAGACTCCATT
Opa-r		CACAGTCGGTCCATTGTTTG
Oligo for chromatin FISH		5×(AAGAG)

2.2 Methods

2.2.1 Embryo collection and fixation

Wild-type embryos (Oregon R) were collected from six population cages (28×17×17 cm, containing 10,000-12,000 flies each, maintained in fly incubators at 25°C and 60% humidity) on 15cm apple juice plates with yeast paste after pre-clearing. The collection windows were 0-4 h after egg deposition (AED) for immunostainings, and 1-2 h, 2-3h, 2-4h, 6-8 h AED for ChIP. After 1-2 min dechorionation in 50% bleach (Clorox), embryos from each plate (~500 mg) were rinsed with deionized water, transferred to 15 ml Falcon tubes containing 2.5 ml Crosslinking Buffer with 1.8% fresh formaldehyde and 7.5 ml heptane, and vortexed at medium speed for 15 min (Embryos collected for FISH were crosslinked 30 min with 4% formaldehyde). After centrifugation at 700 g for 1 min at 4°C, the water and heptane phases were removed and the fixation was stopped by adding 10 ml PBT (1×PBS with 0.1% Triton) with 125 mM glycine for 1 min. After removing glycine PBT, the embryos were vigorously shaken for 1 min in methanol/heptane (50:50) for the devitellinization. After washing three times with methanol, embryos were kept at -20°C in methanol for up to 3 months until needed.

2.2.2 Immunostaining

Embryos stored in methanol were transferred to a 1.5ml microfuge tube and rehydrated at room temperature (RT) through a methanol-PBT gradient (90% Methanol-PBT, 75% Methanol-PBT, 50% Methanol-PBT, 25% Methanol-PBT, PBT, 1 min for each). After washing twice with PBT, embryos were incubated in blocking solution (PBT with 0.5% BSA) for two hours at RT on a rotator. Embryos were then separated into several tubes and incubated with different primary antibodies in blocking solution overnight at 4°C on a rotator. After removing the primary antibodies and washing twice with blocking solution at RT (5 min and 30 min) on a rotator, the embryos were incubated with secondary antibodies (with Alexafluor Dye) in blocking solution for 2 hours at RT in the dark. Embryos were then washed three times in PBT for 10 min and stained with DAPI in PBS. Embryos were rinsed twice in PBS quickly, and about 50 µl of PBS was used to put the tissues on the slide. After removing as much PBS as possible from around the tissues, they were mounted with about 40 µl of VECTASHIELD mounting medium overnight at 4°C before acquisition for confocal imaging (Zeiss LSM-510-VIS). (Mounted tissues can be stored for up to 3 weeks at 4°C.)

2.2.3 Embryo staging

For embryo sorting, wild-type embryos collected within 1-2 h (pre-MBT) and 2-3 h (MBT) were rehydrated as for immunostainings and stained with DAPI. After washing with PBT for 5 min, embryos were sorted in PBT on ice under an inverted contrasting microscope (Leica DMIL). All embryos were screened once for morphology with a DIC filter, and twice for DNA content under UV light (Figure 2-1). Out-of-stage embryos were removed with a 10 µl pipette tip connected to a Cell Tram Vario (Eppendorf: 920002111). After practice, 200 µl embryos (5000 embryos) could be screened in one day.

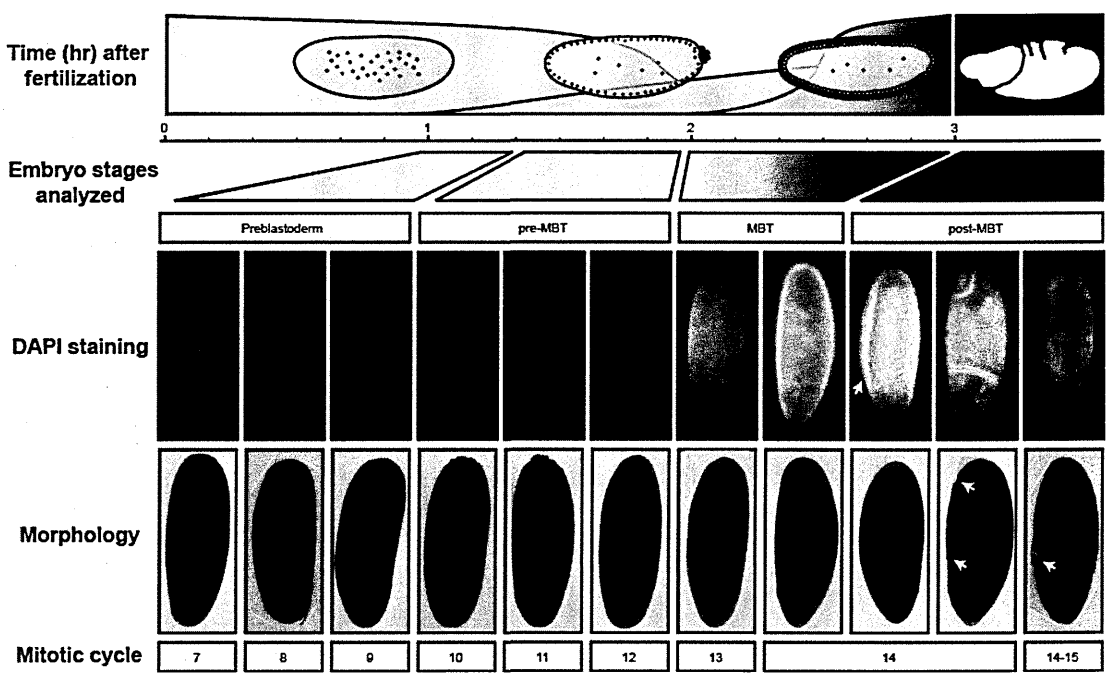


Figure 2-1. Standards for staging pre-MBT and MBT embryos.
Embryos from a pre-MBT collection (1-2 h) were screened for embryos in mitotic cycles 13 and 14 and removed. Using DIC, these embryos were recognized by their cellularization or gastrulation furrows (marked by arrows in the post-MBT embryos), and were removed. Under UV light, these embryos were recognized based on the number of nuclei visible using DAPI staining. Likewise, for MBT collections (2-3 h), embryos with visible furrow formations or high DAPI staining signals were removed. The small number of embryos that had fewer than the desired number of mitotic cycles were tolerated in both pre-MBT and MBT collections.

2.2.4 Chromatin Immuno-Precipitation (ChIP)

ChIPs were performed as previously described (Zeitlinger et al., 2007) using 200 µl embryos for pre-MBT embryos and 50 µl for older embryos (the amount of embryos used for pre-MBT and MBT ChIP was calculated based on nuclear numbers, that is, the number of nuclei in MBT embryos is around 4 fold higher than in pre-MBT embryos). 0.5 g embryos was transferred to a 2 ml Dounce homogenizer with 5 ml Buffer A1 with fresh 1

× Protease Inhibitor Cocktail and homogenized five times each with the loose fitting and the tight fitting pestles. Homogenized samples were transferred to 15 ml falcon tube and then centrifuged 3 min at 1500 g. The pellet was subsequently washed three times with Buffer A1 and once with Buffer A2 containing 1 × fresh Protease Inhibitor Cocktail by first resuspending the pellet with a serological pipette and centrifuging in-between washes. Finally, all remaining buffer was removed, the pellet was resuspended in 1.2 mL Buffer A2, and 300 µL aliquots were transferred to fresh 1.5 mL microfuge tubes. The homogenate was then sonicated in a Bioruptor waterbath sonicator for 15 cycles (30 s on/off, "high" setting). After sonication the tubes were centrifuged 10 min at 4°C at maximum speed to pellet debris and insoluble chromatin. The supernatants containing soluble chromatin were pooled for ChIP. The total volume was adjusted to 700 µl with buffer A2 when necessary. 50 µl chromatin was set aside as input. 10 µg antibodies was added and incubated overnight at 4 °C with rotation.

ChIPed chromatin was purified by adding 50 µl pre-washed Dynabeads coated with antibodies against rabbit or mouse IgG for 4 h rotating at 4 °C. The beads were washed 3 times in RIPA buffer and once in TE + salt. Immunoprecipitated DNA was eluted from the beads twice in 100 µl elution buffer at 65°C to maximize yields. TE was added to 400 µl volume, then crosslinks of ChIP and input DNA were reversed over night at 65 °C. The DNA was purified by RNase A (60 µg, 30 min at 37 °C) and proteinase K (60 µg, 2 h at 56 °C) treatment followed by phenol/phenol-chloroform-isoamylalcohol extractions and ethanol precipitation. The precipitated DNA was resuspended in 35 µl H_2O. 5 µl was set aside to evaluate the ChIP enrichment by real-time PCR (StepOnePlus; Fast SYBR Green Master Mix; Applied Biosystems). The remaining 30 µl ChIP DNA was used to construct ChIP-Seq libraries, with 100 ng input DNA as control.

2.2.5 Library preparation of ChIP DNA for Illumina sequencing

Libraries were prepared from the immunoprecipitated DNAs using the Paired-End DNA Sample Preparation Kit but with a modified protocol. To remove adapter dimers, Biotin-14-dATP was added in the A-tailing reaction after end-repair. After ligation to the PE adaptor, the samples were incubated with streptavidin beads in 500 µl Binding & Washing buffer (B&W buffer) at RT for 15 min. DNA bound to the beads was then washed twice with 800 µl B&W buffer with 0.05% Tween 20, twice with NEB buffer 2, and resuspended in 31 µl NEB buffer 2. The PCR reaction was then performed according to the Illumina protocol.

2.2.6 MNase-Seq

50 µl sorted 2-3 h embryos was homogenized as ChIP-Seq and digested based on a previously published protocol (Mavrich et al., 2008). Briefly, homogenized chromatin in 4.5 ml NPS buffer was aliquoted into nine microfuge tubes and digested with a Micrococcal nuclease (MNase) gradient of 20 U, 10 U, 5 U, 5/2 U, 5/4 U, 5/8 U, 5/16 U, to 5/32 U, and one negative control for 30 min at 37°C. Digestion was stopped by adding 20 µl 0.5M EDTA to each tube and keeping on ice for 10 min. After treating with RNase and proteinase K, reverse crosslinking was performed at 60°C for 6 h, and the DNA was cleaned up as described in ChIP-Seq. The ethanol-precipitated DNA was resuspended in 10 µl ddH_2O and run on a 1.7% agarose gel (See Figure 3-14). Mono-nucleosome sized DNA was extracted from the lane with two clear bands, and prepared for paired-end sequencing.

2.2.7 Probe preparation for chromatin Fluorescence *in situ* hybridization (FISH)

pBac DNA was purified following the protocol from Children's Hospital Oakland Research Institute. A single bacterial colony of each BAC clone was transferred into 3ml LB media with 25 µg/ml kanamycin. After growing overnight (up to 16h) at 37°C 225-300 rpm, the media was centrifuged at 12,000 g for 30 s at 4°C, and supernatants were discarded. Each pellet was resuspended in 300 µl P1 solution. After adding 300 µl of P2 solution, contents were gently shaken until mixed, and then kept at RT for 5 min. 300 µl P3 solution was slowly added to each tube with gentle shaking, then the tubes were placed on ice for 15 min. After spinning at 12,000 g for 10 min at 4°C, supernatant was transferred to a tube with an equal volume of ice-cold isopropanol and mixed by inverting the tube a few times. Tubes were spun at 12,000 g for 20 min at 4°C and then the supernatant was removed. DNA pellets were washed with 0.5 ml of 70% EtOH twice, and then spun at 12,000 g for 5 min at 4°C. As much of the supernatant as possible was removed and then pellets were air-dried at RT. DNA was resuspended in 40 µl TE with 1 µl RNase and the identity of the BAC was confirmed by sequencing.

BAC plasmid was sonicated in 300 µl ddH_2O for 5 min at high power and BAC fragment was purified with Promega PCR clean-up kit and eluted with 300 µl ddH_2O. 20 µg BAC fragment was denatured in an EP tube with 50 µl H_2O at 95°C for 5 min and chilled immediately on ice. (For the synthetic oligo against pericentromeric region, skip the denaturing step.). Denatured DNA was add to the following reagents at 4°C:

20 µl of 5×Tdt buffer

20 µl of 25mM $CoCl_2$
3 µl 2mM amine-dUTP
6 µl 1mM dTTP
1 µl (400 units) TdT
Incubate at 37°C for 2h;

Reaction was stopped by adding 1 µl 0.5M EDTA, and DNA was purified using MinElute PCR Purification Kit with 10 µl H_2O. 6 µl of kit labeling buffer was added to 20 µg probes. Dye was dissolved in 4 µl DMSO and added to DNA. The labeling reaction was incubated in the dark for 1-2 h at RT, and then DNA was purified by ethanol precipitation and air-dried. The labeled DNA was re-suspended the in 200 µl FHB to adjust the final concentration to be 100 ng/µl.

2.2.8 *in situ* hybridization

Fixed embryos were re-hydrated into PBT as described above and incubated in 1ml PBT with 20 µl 20 mg/ml RNase for 2 h at RT on a rotator. After another 1 h incubation at RT in 1 ml of PBT, embryos were transferred into a pre-Hybridization Mixture (pHM) by passing through the following PBT (PBS with 1% Triton 100)- pHM gradient (1 ml and 20 min each step): 80% PBT / 20% pHM, 50% PBT / 50% pHM, 20% PBT / 80% pHM, 100% pHM. Embryonic DNA was denatured by incubation for 15 min at 80°C in 100% pHM. After 5 min, the probes in FHB were denatured by incubating for 10 min at 95°C. While embryos were still at 80°C, buffer was removed, and the denatured probes in FHB were added to the tissues without prior cooling. The tissue suspensions were covered with a drop of mineral oil and then incubated in a thermomixer at 37°C with 450 rpm agitation for 14-17h.

Embryos were then washed with wash gradient [50% formamide in 2X SSC with 0.3% CHAPS, 40% formamide in 2X SSC with 0.3% CHAPS, 30% formamide / 70% PBTw (PBS with 0.1% Tween 20), 20% formamide / 80% PBTw, 10% formamide / 90% PBTw] in the thermomixer for 20 min with 800 rpm agitation at 37°C (1ml each step). Embryos were then washed with PBTw and PBT for another 20 min each on a rotator at RT, and then were stained with DAPI for another 15 min. Finally, embryos were mounted as described for immunostaining.

2.2.9 Sample preparation for Electron Microscope (EM)

Wildtype embryos collected within 0-4 h were dechorionated as described above and were fixed following previous protocol (McDonald et al., 2012). PBT was carefully removed and then replaced with heptane saturated with glutaraldehyde (combined 2 ml of Fixative A with 8 ml of heptane and shook vigorously, allowed phases to separate, and used the upper phase for fixation) at RT for 25 min. Then fixative was removed with a cut-off yellow micropipette tip and embryos were transferred to the siliconized glass slide. After the heptane evaporated, a piece of double-coated adhesive tape was carefully pressed to embryos and transferred to a 35-mm petri plate with the tape with embryos up in Fixative B. Next, embryos were dissected out of their vitelline membrane under a dissecting microscope with #5 forceps (FST), and then transferred to Fixative C on ice for an additional 2 h. embryos were washed twice with 50 mM sodium cacodylate, and then post-fixed in 1% osmium tetroxide on ice for 2 h before submission to EM facility.

CHAPTER 3

Results

3.1 Massive *de novo* recruitment of Pol II during MBT

3.1.1 Markers of Pol II recruitment are gradually accumulated during MBT

To understand when Pol II is recruited during early embryogenesis, I first examined markers of transcription initiation in *Drosophila* embryos by immunostaining. As shown in Fig 3-1, Pol II can be detected in the nuclei of embryos as early as the pre-blastoderm stage (cycles 2-8), during which few zygotic genes are transcribed (Zalokar, 1976). This suggests that Pol II can be maternally deposited as protein and transported into nuclei before ZGA. However, there is no signal of serine-5-phosphorylated (Ser5pho) Pol II, a marker for Pol II initiation, in the nuclei until cycle 9, which is consistent with previously reported observations of early zygotic transcription (Lamb and Laird, 1976). The signal of Ser5pho Pol II increases during cycle 14 when the zygotic genome is globally activated (Anderson and Lengyel, 1980). This signal increase suggests that transcription initiation starts with a small portion of genes in the pre-MBT embryos, but for the vast majority of transcribed genes, transcription begins during MBT.

I further analyzed the general pattern of TBP during embryogenesis by immunostaining. I found that TBP can only be detected in the nuclei after cell cycle 8 and increases during MBT, similar to the appearance of Ser5pho Pol II. Since TBP mRNA is maternally deposited, this indicates that the appearance of TBP protein in the nuclei is regulated at the translational or post-translational (such as importing into nuclei) level in the early embryo. Taken together, my immunostaining results suggest that Pol II recruitment may start after cell cycle 8, when a few pre-MBT genes are beginning to be transcribed, and massive Pol II recruitment takes place during MBT.

3.1.2 Global *de novo* Pol II recruitment during MBT

Next, I performed ChIP-Seq experiments to analyze the exact occupancy of Pol II, TBP, and histone modifications in pre-MBT embryos (mitotic cycles 10-12), MBT embryos (mitotic cycles 13-14), as well as post-MBT embryos as a control. Although the large quantity of *Drosophila* embryos required for ChIP-Seq can be collected by conventional means, such collections always contain a fraction (5-20%) of older embryos due to maternal egg holding. I tested different methods for eliminating this contamination and ultimately decided to stain embryo collections with DAPI and remove "out-of-stage" embryos under a microscope with a pipette (Methods of Chapter 2, Figure 2-1, and Figure

3-2a). In addition, I adjusted the ChIP protocol to accommodate the use of fewer embryos to make this experiment more feasible (Methods of Chapter 2). The ChIP-Seq data from these hand-sorted embryos have robust and reproducible signals in replicates (Fig 3-3a, b).

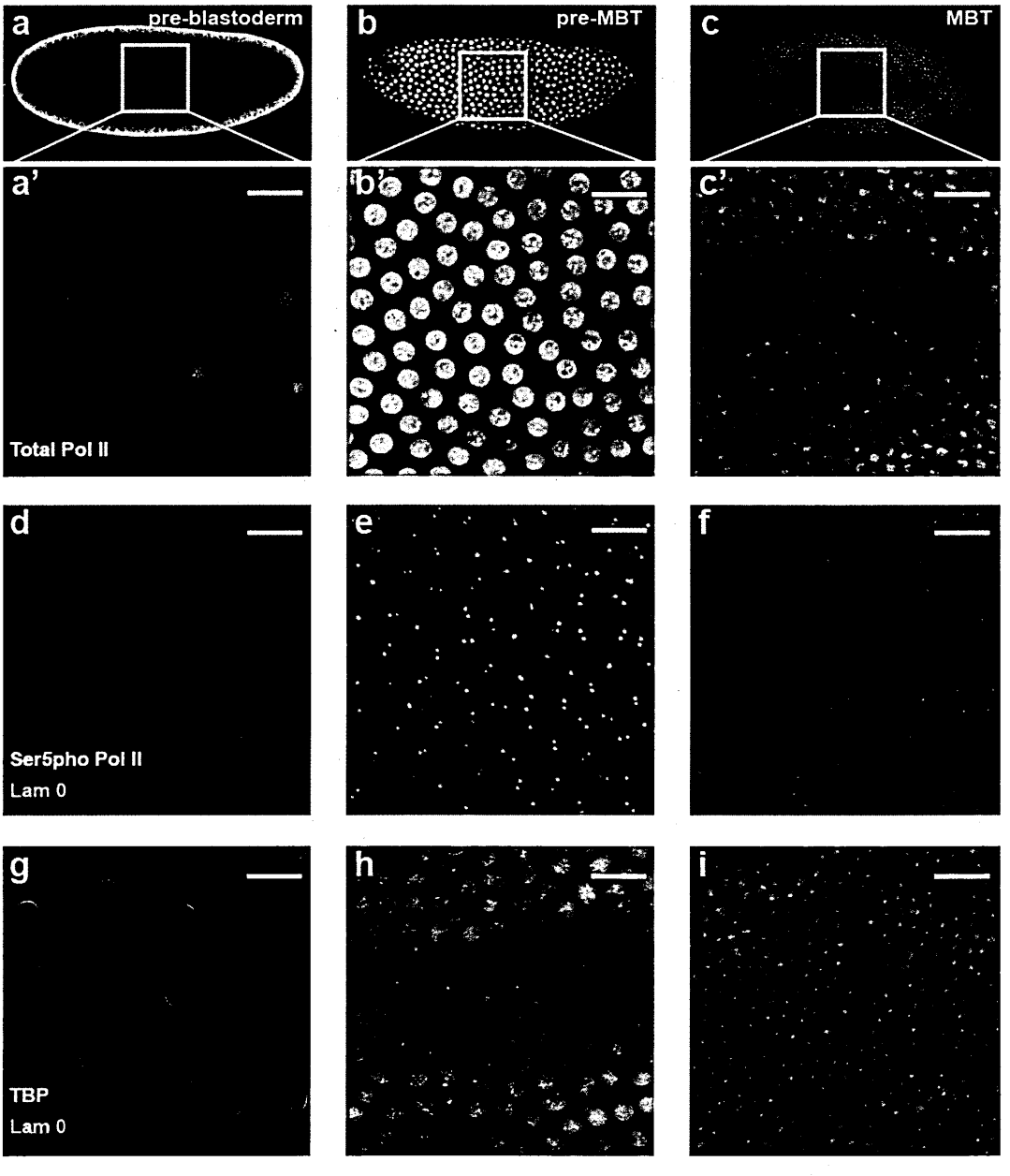


Figure 3-1. Markers of Pol II recruitment are gradually accumulated during MBT.
a, b, c, Immunostainings of unphosphorylated Pol II suggest that Pol II is deposited as protein by maternal deposit in pre-blastoderm embryos (mitotic cycles 1-7). However, the initiated form of Pol II (serine-5-phosphorylation of the CTD repeats. **d, e, f**), as well as TBP (**g, h, i**) are only detectible in the nuclei (outlined by the lam 0 in red) of pre-MBT embryos (mitotic cycles 8-12) when zygotic transcription begins (scale = 20μm).

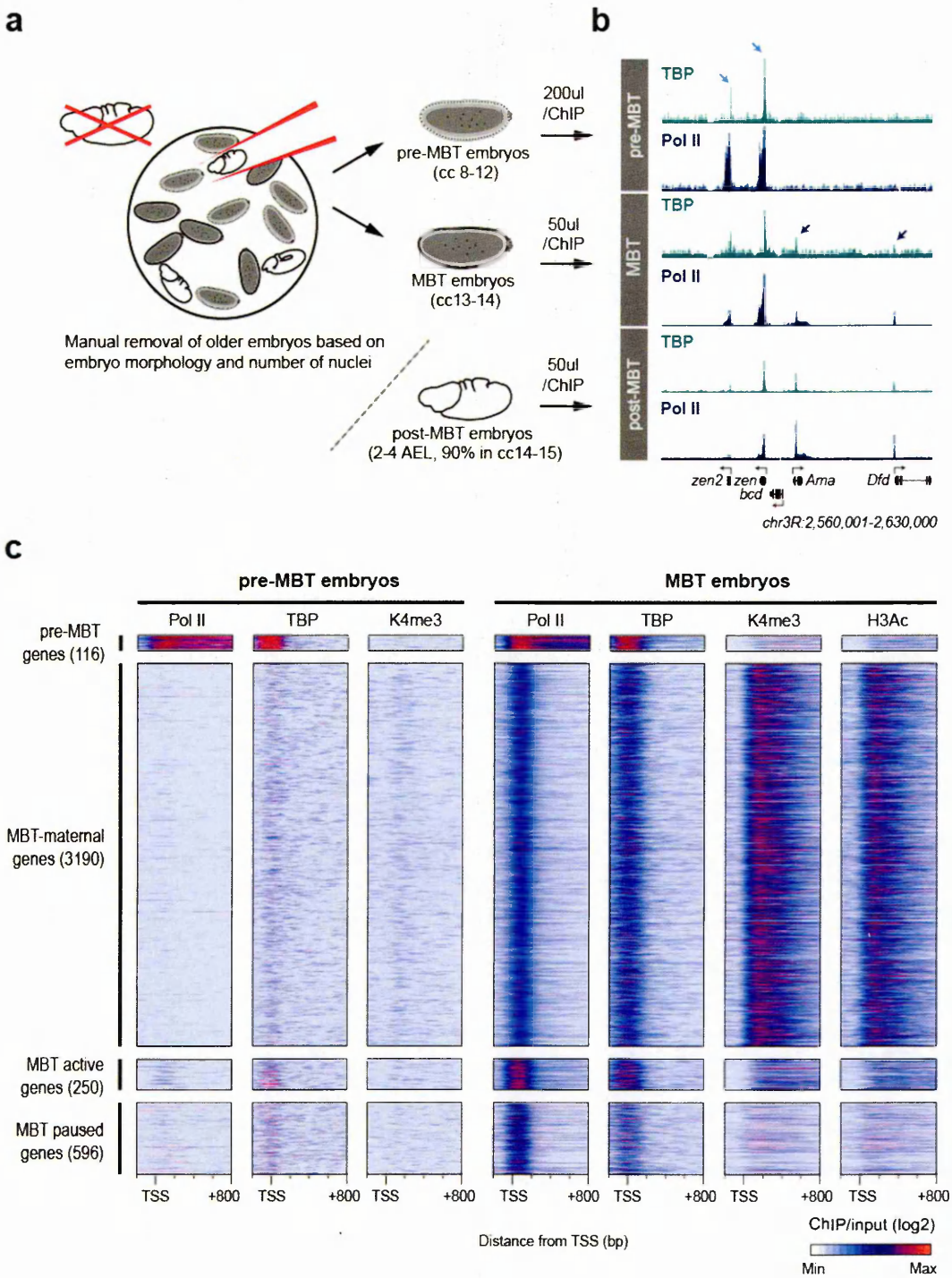


Figure 3-2. Global recruitment of Pol II during MBT.
a, Outline of the hand-sorting of embryo collections for ChIP-Seq experiments. **b,** A snapshot of a genome browser view of Pol II and TBP ChIP-Seq profiles. The sorting purity is reflected by unique peaks in pre-MBT and MBT samples **c,** Heat map of ChIP-Seq enrichments across all genes (aligned at TSS) that are significantly bound by Pol II during MBT. Note the massive *de novo* recruitment of Pol II during MBT.

In the pre-MBT embryos, both Pol II and TBP are only found in a small number of genes, including most of the pre-MBT genes previously defined by *in situ* hybridization (Summarized in Bosch et al., 2006) or microarray data (De Renzis et al., 2007) (Figure 3-

3c), as well as core histones (but not the linker histone) in the histone repeat cluster (Anderson and Lengyel, 1980). In contrast, there is no obvious Pol II or TBP occupancy at most housekeeping genes or developmental genes, which are transcribed or paused after pre-MBT. This is consistent with the low staining signals of Ser5pho Pol II and TBP in the pre-MBT embryos, and suggests that there is no pre-recruited Pol II or TBP in the pre-MBT embryos for preparing massive ZGA during MBT. Indeed, I did observe *de novo* recruitment of Pol II and TBP at the promoters of ≈ 4,000 genes during MBT, which equates to roughly a third of all *Drosophila* genes (Figure 3-2c).

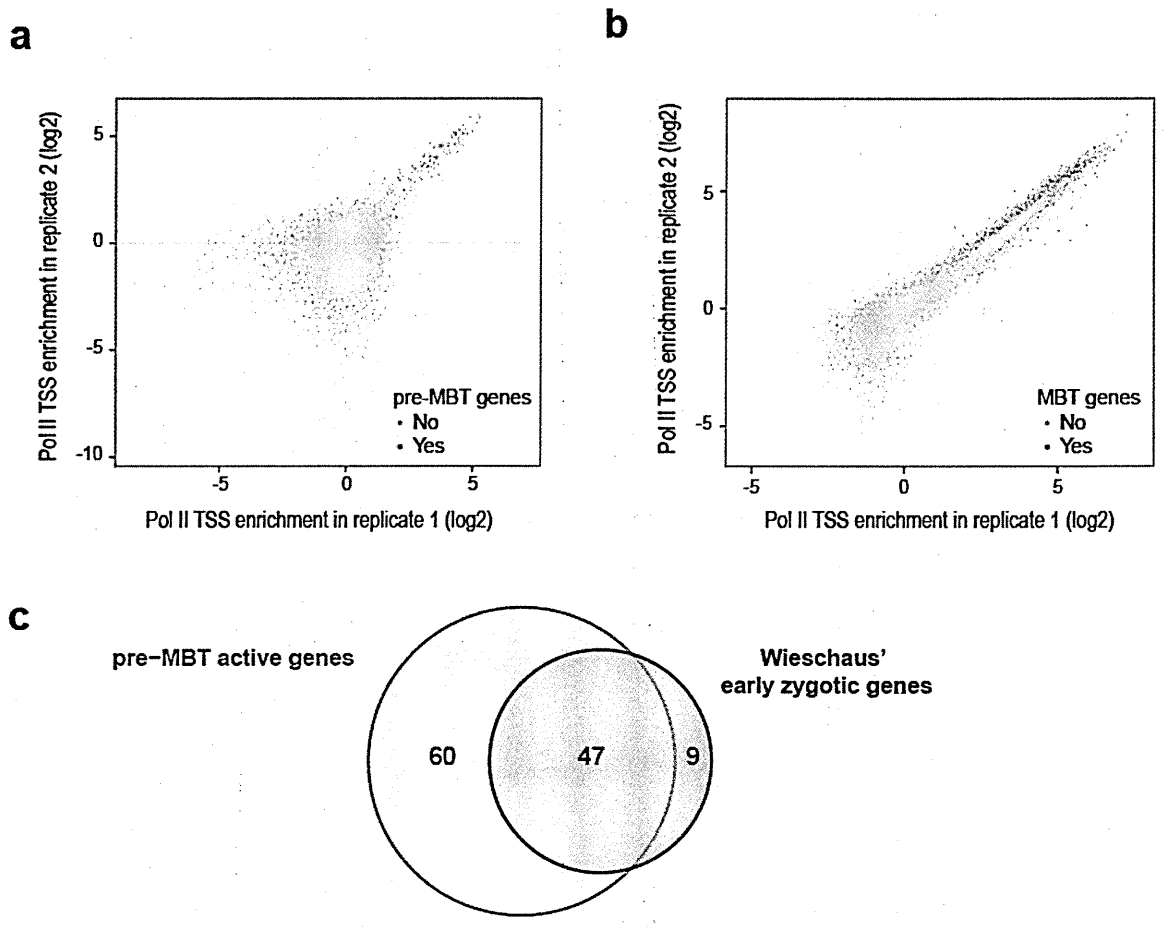


Figure 3-3. Reproducibility of Pol II ChIP-Seq between biological replicates.
a, Scatterplot comparing two pre-MBT Pol II replicates. For each replicate, the enrichment over background in each TSS region (first 200bp) is displayed. Genes that qualified as pre-MBT genes based on four replicates are shown in red. **b,** The same scatterplot comparing two MBT Pol II replicates. Genes that were classified as MBT genes are colored in red. **c,** Venn diagram showing the strong overlap between active pre-MBT genes (those in the "paused later" and "never paused" group, without the false positive ones) and the early zygotic genes identified previously (De Renzis et al., 2007).

3.1.3 No apparent Pol II pausing at the earliest transcribed genes

To obtain a confident list of single-copy genes occupied by Pol II before MBT ("pre-MBT genes"), Jeff and I first identified all genes with at least two-fold enrichment of Pol II over input at the transcription start site across all four Pol II ChIP-Seq replicate experiments in pre-MBT embryos. From this list, we removed 12 genes that were

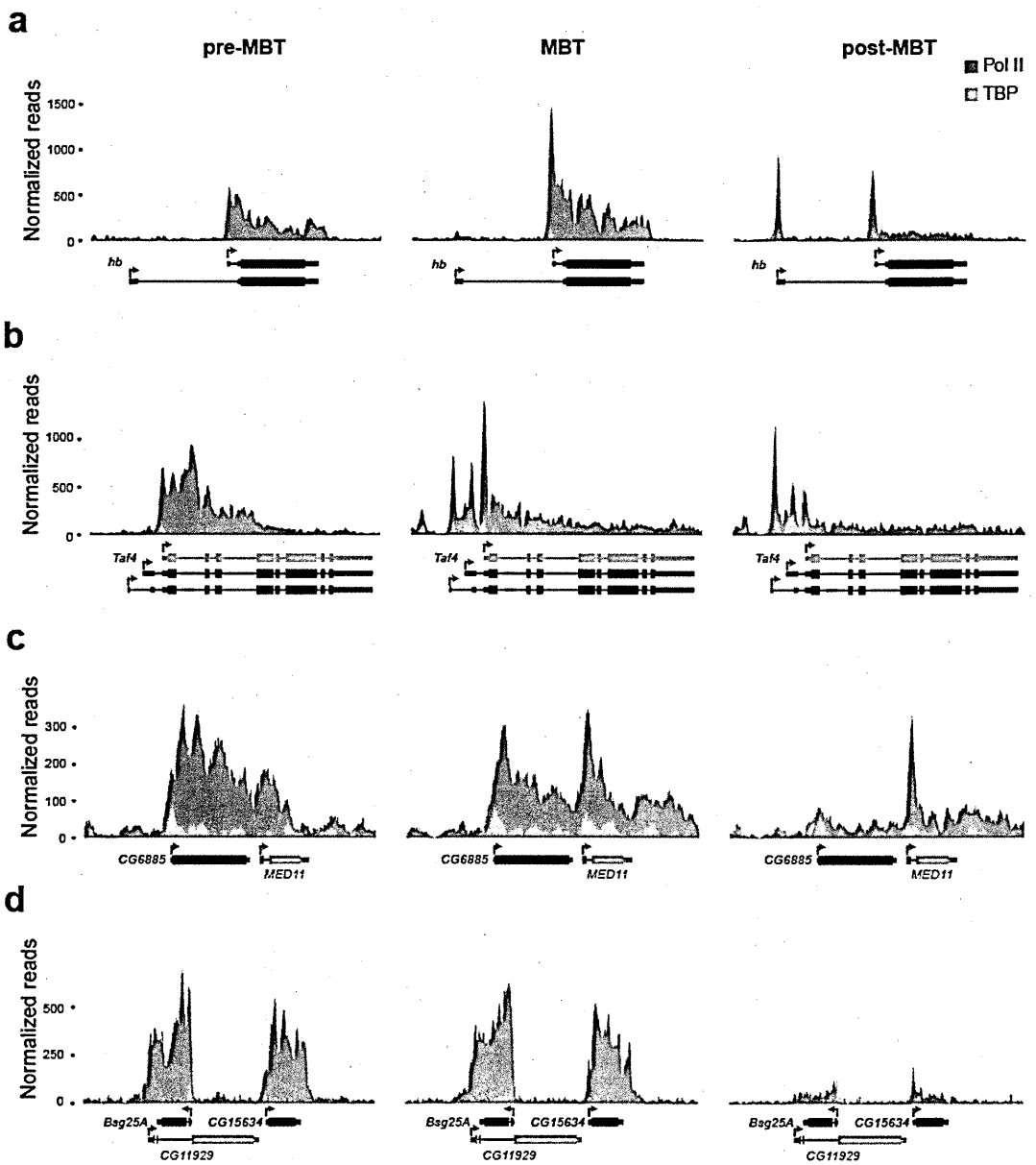


Figure 3-4. ChIP-Seq occupancy of Pol II and TBP at genes with complex patterns.
a, *hb,* a pre-MBT gene with known alternative TSSs. While the most proximal TSS is preferentially used in pre-MBT and MBT embryos, the distal TSS is used mostly in the post-MBT stage. **b,** *Taf4,* a gene with an un-annotated, more proximal TSS that is used during the pre-MBT stage (light grey). The two distal, known TSSs are used during MBT. 22 out of 35 pre-MBT genes use the proximal TSS during early embryogenesis. **c,** *MED11,* shown in red, an example of a false pre-MBT gene due to read-through signal from an upstream gene, *CG6885.* **d,** *CG11929,* shown in red, an example of a false pre-MBT gene due to signal from an overlapping gene, *Bsg25A (detail in the Appendix).*

likely false positives as a result of Pol II read-through from a nearby gene and added 10 genes that were missed due to un-annotated alternative start sites (see examples in Figure 3-4 and Methods in Appendix). This yielded 116 single-copy pre-MBT genes, many of which have known functions in early AP patterning, DV patterning, sex determination, and cellularization (Table 3-1), while 43 have unknown functions. Consistent with previous study, most of these pre-MBT genes are short (<2,000 bp) and intron-less.

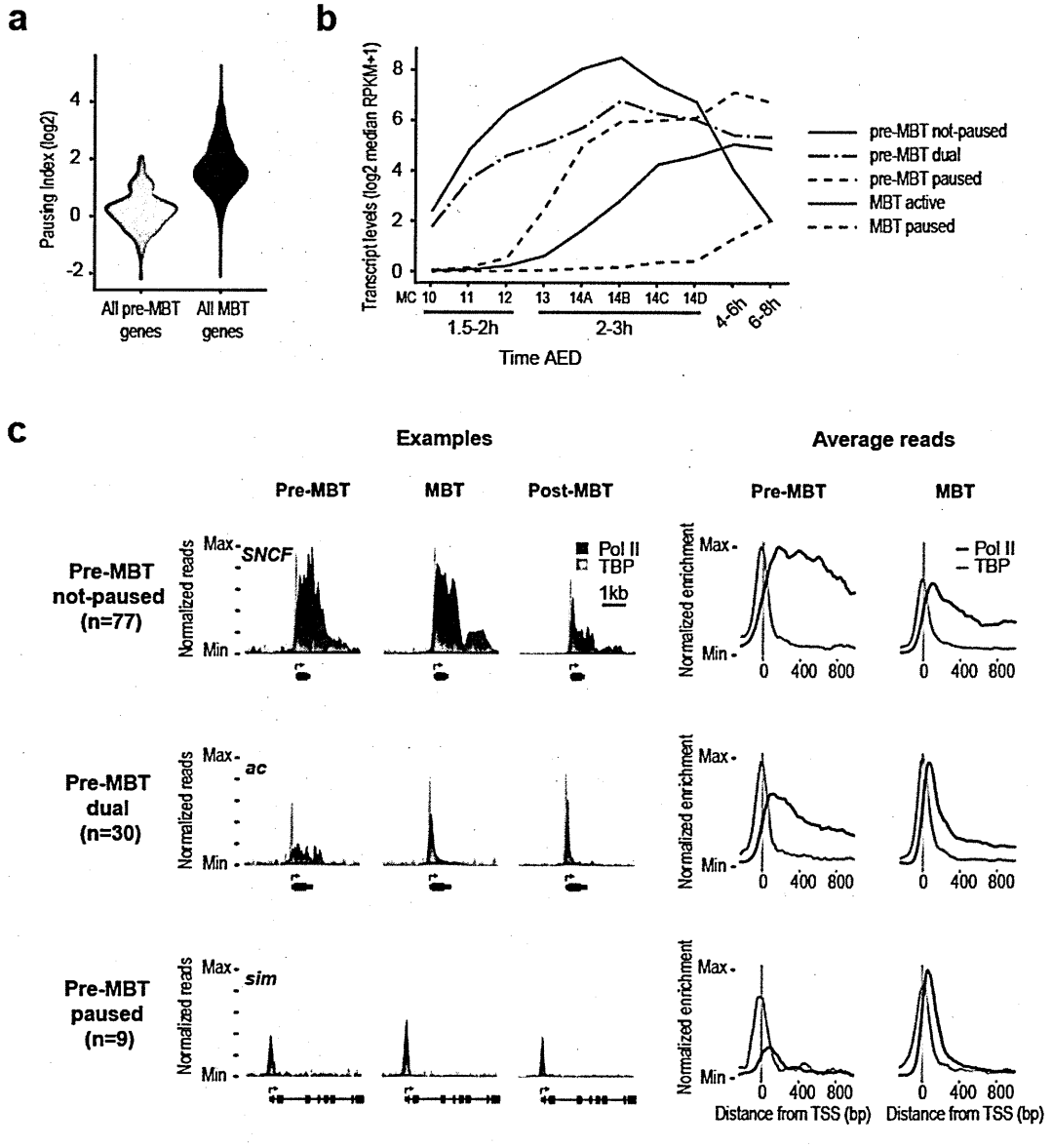


Figure 3-5. Lack of Pol II pausing during pre-MBT.
a, A violin plot of the Pol II pausing index distribution shows that pre-MBT genes (during pre-MBT stages) display less Pol II pausing than MBT genes (during the MBT stage). **b,** Median RNA-Seq expression data (Lott et al., 2011) of the three pre-MBT groups and the two MBT groups show that paused genes are expressed at lower levels and may be induced later. **c,** Examples and average enrichment of Pol II occupancy (blue) and TBP occupancy (pink) for the three pre-MBT gene groups.

Inspection of Pol II occupancy revealed that most pre-MBT genes have Pol II across the whole gene body without any notable peak at the pause site [+50 from the transcription start site, (Zeitlinger et al., 2007)] when they are initially transcribed, while TBP occupancy is found at the expected position (~30 bp from the transcription start site) (Figure3-4, 3-5c). When quantifying the degree of pausing using the pausing index (Pol II signal at TSS/TU, see Methods in Appendix), pre-MBT genes are indeed much less paused than MBT genes (Fig. 3-5a). However, a small number of pre-MBT genes have higher pausing indices and lack Pol II enrichment within the gene body and thus may be paused.

Table 3-1. Classification of all active pre-MBT genes

Functions	Gene names
Sex determination	**Dpn**, sisA, Sxl, os
Cellularization	nullo, Sry-alpha, kuk, bnk, slam
Anterio-posterior patterning	**cad**, **hb**, **gt**, **kni**, **tll**, **eve**, **h**, **run**, **slp1**, **odd**, ftz
Dorso-ventral patterning	**sna**, **esg**, **Nrt**, **glec**, **ac**, **l(1)sc**, **Tom**, **BobA**, **m4**, zen, zen2, tsg, tld, scw, Neu2, sc, fd19B, bnb, Bro, Brd, Ocho, amos, ato
Other function	**Taf4**, **wech**, **Corp**, **toc**, **spri**, Z600, halo, SNCF, CG4570, spo, hrg
Non-coding RNA	mir-9a, mir-309, roX1, snRNA:U5:34A, snRNA:U4atac:82E, snRNA:U1:82Eb, snRNA:U5:23D, snRNA:U5:38ABb, snRNA:U5:14B, snRNA:U4:38AB, snRNA:U1:95Cc
Unknown function	
Localized expression	**gk**, **CG9894**, **CG5059**, sala, term, CG14427, CG8960, CG13711, CG13713, CG15876, CG6885, CG7271, CG14014
Ubiquitous expression	Bsg25A, Bsg25D, CG15634, CG15382
Others	**CG2201**, **CG42666**, CG43659, CG13716, CG13712, CG13000, CG13465, CG14561, CG18269, CG14915, CG16813, CG15479, CG15480, CG4440, CG14317, CG13427, CG34137, CG34214, CG34224, CG34266, CG16815, CG42762, CG43184, CG 9775, CG9883, CR43887

Note: bold marks the pre-MBT dual genes

I therefore examined the pre-MBT genes based on Pol II occupancy and expression levels [RNA-Seq data by (Lott et al., 2011)] and identified three distinct groups (Figure 3-5b, c). Genes in the first group ("pre-MBT not-paused", n=77) have highest expression during cellularization (cycle 14) and tend to diminish in expression thereafter (Figure 3-5b). These genes appear never to become paused during early development (see *SNCF* and the average profile in Figure 3-5c). Genes in the second group ("pre-MBT dual", n=30) initially show little evidence of pausing; however, Pol II gradually accumulates at the pause site during and after MBT (see *ac* and the average profile in Figure 3-5c). Finally, there is a small group of genes ("pre-MBT paused", n=9) that appear to have paused or non-productive Pol II even at the pre-MBT stages (see *sim* and the average profile in

Figure 3-5c), which is consistent with the expression of these genes. Their transcript levels rise much later during pre-MBT stages compared to the first two groups (Figure 3-5b). This suggests that Pol II pausing can exist during pre-MBT stages, but most pre-MBT genes—even those that are eventually paused—are non-paused during early pre-MBT.

3.1.4 Widespread Pol II pausing is established during MBT for later activation.

I next analyzed Pol II patterns at genes newly occupied by Pol II during MBT ("MBT genes"), which includes most housekeeping genes and many developmental genes. Jeff and I first subtracted from them the 3,190 maternal genes ("MBT-maternal genes"), which have high RNA-Seq reads in cycle 10 (Lott et al., 2011) when maternal RNA degradation just starts, and they are also known to be enriched for broadly expressed housekeeping genes (Rach et al., 2009). The remaining 846 genes ("MBT-zygotic genes") frequently have high Pol II occupancy at the pausing site and a high pausing index (Figure 3-5a), suggesting that Pol II pausing is widespread during MBT.

Although the widespread Pol II pausing is coincident with the massive ZGA, it is not just associated with activated genes. Among MBT-zygotic genes, only 250 (30%) of these genes are expressed at significant levels during late cycle 14 ("MBT active genes", see *brk* and the average profile in Figure 3-6a). The remaining 596 genes ("MBT paused genes", see *Dr* and the average profile in Figure 3-6a) are expressed at very low levels, which is typical for paused genes (Zeitlinger et al., 2007).

This raises the possibility that paused Pol II at quiescent genes may prepare them for later activation. To test this hypothesis, Jeff and I analyzed expression patterns of these two groups MBT-zygotic genes using the large-scale *in situ* hybridization database ImaGO [http://insitu.fruitfly.org/ (Tomancak et al., 2002; 2007)]. As expected, the MBT active genes start to be transcribed very early, and most are first detected at stages 4-6 (MBT), when MBT occurs. In contrast, the MBT paused genes tend to be first detected at later embryonic stages (stages 9-10 or later, after gastrulation) (Figure 3-6b). Thus, many newly transcribed genes become paused during MBT for later activation.

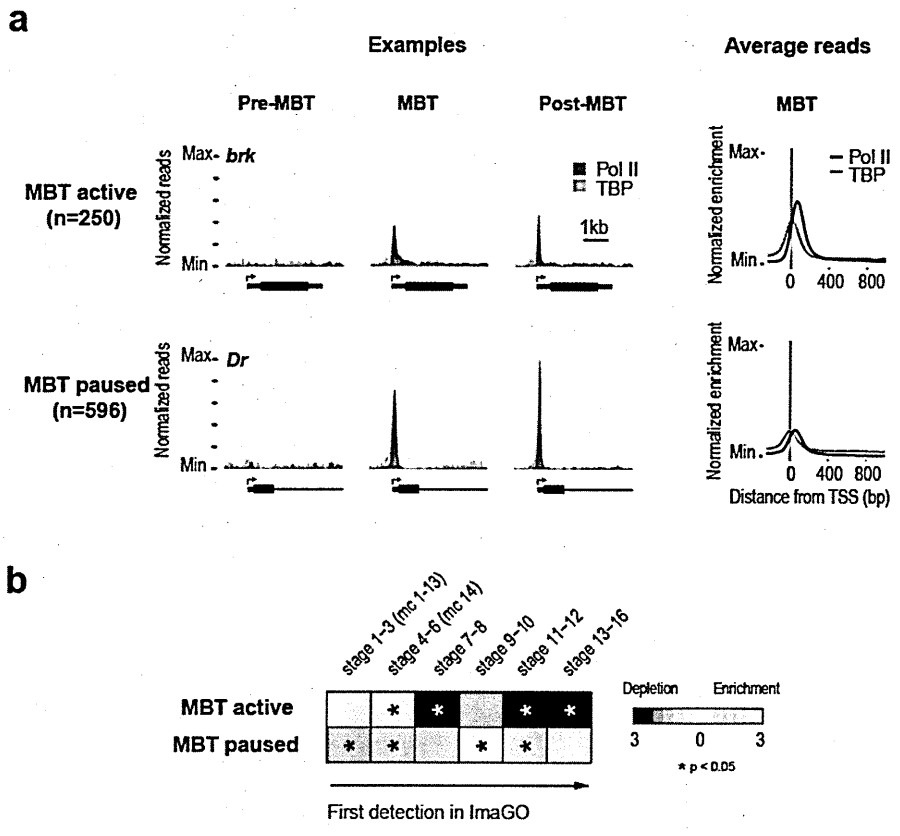


Figure 3-6. Widespread Pol II pausing during MBT.
a, Examples and average enrichment of Pol II occupancy (blue) and TBP occupancy (pink) for the two MBT gene groups based on single embryo RNA-Seq data **(Lott et al., 2011). b,** Analysis of large scale *in situ* hybridizations (ImaGO database, see Methods) confirms the earlier initial expression (first detection) of MBT active genes (mostly stage 4-6, around MBT) compared to MBT paused genes (mostly stage 9-10, after gastrulation).

3.2 No bivalent domain for instructing the onset of ZGA

3.2.1 Gradual deposition of histone modifications associated with transcription during MBT

To test whether histone modifications may be present before gene activation during early embryogenesis, I first performed immunostaining in the early embryos with antibodies against H3K4me3 and H3K9/14ac. These histone modifications are associated with gene activation, but are also found at quiescent developmental genes in mammalian ES cells (Guenther et al., 2007). It has been recently reported that H3K4me3 is present in early zebrafish embryos before the ZGA (Lindeman et al., 2011). However, I could not detect H3K4me3 in nuclei from pre-blastoderm embryos prior to gene activation. Apart from this, there is no detectable H3K4me3 in pre-MBT embryos while ~100 pre-MBT genes are actively transcribed. H3K4me3 begins to be stably detected in the nuclei during cellularization, when the bulk of zygotic transcription starts, and increases during embryogenesis.

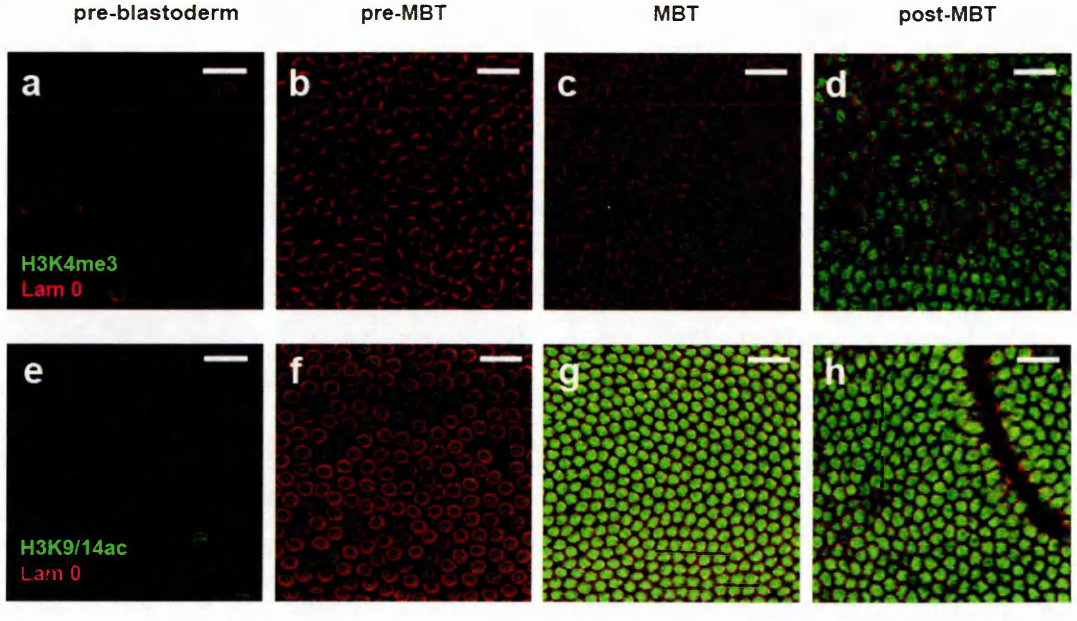


Figure 3-7. Differential deposition of histone modifications associated with euchromatin during development.
a,b,c,d, Immunostainings of H3K4me3 suggest that H3K4me3 is modified *de novo* during MBT. In contrast, H3K9/14ac can be detected in pre-blastoderm embryos, when there is little zygotic transcription. The signal of H3K9/14ac increases during embryogenesis (scale = 20μm).

In contrast to H3K4me3, H3K9/14ac is visible in nuclei from early embryos prior the pre-MBT gene activation. Meanwhile, it is also detected in polar bodies, the siblings of the maternal pre-nucleus, in pre-MBT embryos. The staining signal of H3K9/14ac increases significantly during ZGA, and it continues to intensify during later development. This raises the possibility that H3K9/14ac, unlike H3K4me3, may be maternally inherited and could in principle be a pre-patterning marker if such marks exist in *Drosophila* embryos.

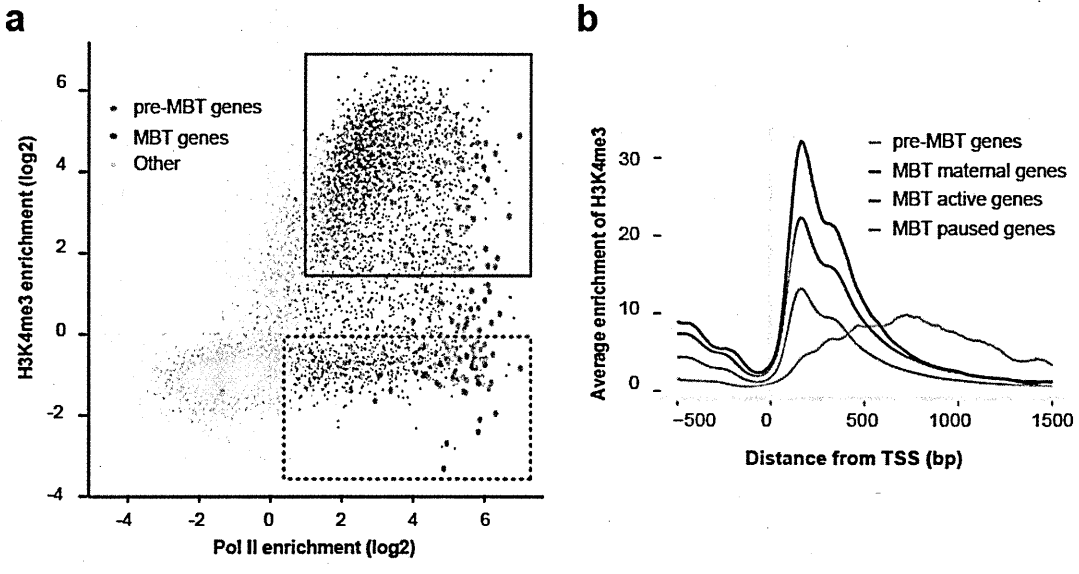


Figure 3-8. H3K4me3 is associated with active transcription in MBT embryos.
a, Scatterplot comparing H3K4me3 and Pol II enrichment in MBT embryos. In general, H3K4me3 is associated with Pol II recruitment (shown in solid line). However, there is a portion of genes (including many pre-MBT genes marked by red dots) that have high Pol II enrichments but little H3K4me3 (shown in dotted line). The grey points to the left of the solid line are due to single ChIP results, which have not been called MBT genes in replicate experiments. **b,** Average enrichments of H3K4me3 in different gene groups show that H3K4me3 downstream of the TSS is highly associated with the transcription status of MBT genes but not pre-MBT genes.

3.2.2 H3K4me3 and H3K9/14ac are associated with active transcription during MBT

To further test whether H3K4me3 and H3K9/14ac might serve as pre-patterning markers for transcription during early embryogenesis, I performed ChIP-Seq experiments in *Drosophila* pre-MBT, MBT, and post-MBT embryos. In general, the ChIP-Seq results for H3K4me3 and H3K9/14ac are consistent with the immunostaining results in early embryos.

There is no obvious signal of H3K4me3 at any TSS region in pre-MBT embryos, not even at pre-MBT active genes. The global deposition of H3K4me3 co-occurs with the

massive Pol II recruitment in MBT embryos. As shown in Figure 3-2c, H3K4me3 is frequently found further downstream of the TSS and is strongly associated with gene transcription levels (Figure 3-8). Genes actively transcribed, like MBT-maternal genes and MBT-active genes, have high H3K4me3 signals, while MBT-paused genes have little signal. This is consistent with our observations in embryonic muscle tissues (Gaertner et al., 2012) that H3K4me3 only associates with active genes rather than paused genes. Intriguingly, most pre-MBT genes have very low levels of H3K4me3 in MBT and post-MBT embryos, even though they are highly transcribed and their Pol II signal is higher than those of MBT-maternal genes. This explains the late appearance of H3K4me3 in immunostaining and ChIP-Seq data, although pre-MBT genes have already begun transcription.

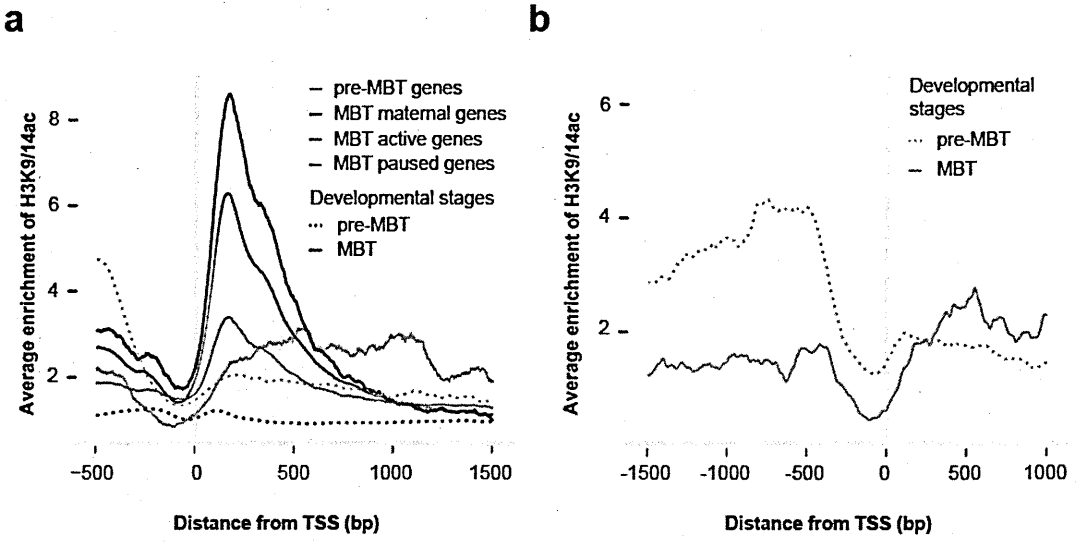


Figure 3-9. Distinct H3K9/14ac patterns in MBT and pre-MBT embryos.
a, Average enrichment of H3K9/14ac in different gene groups in MBT embryos show that H3K9/14ac downstream of the TSS is highly associated with the transcription status of MBT genes, whose H3K9/14ac is absent in the pre-MBT embryos (shown in black dot line). **b,** Comparison of average enrichment of H3K9/14ac in pre-MBT and MBT embryos show a unique pattern of H3K9/14ac upstream of promoters of pre-MBT genes, which decreases during development.

The pattern of H3K9/14ac occupancy is similar to that of H3K4me3 in the MBT and post-MBT embryos: 1) It is deposited further downstream of the TSS and is only associated with the actively transcribed MBT genes in both MBT and post-MBT embryos (Figure 3-8b); and 2) little H3K9/14ac is deposited downstream of most pre-MBT genes (Dot line in Figure 3-9a), despite their active status. Although there is H3K9/14ac staining in pre-MBT embryos, no signal of H3K9/14ac was found associated with any MBT genes,

neither active nor paused, in the pre-MBT embryos. However, at 500-1,500 bp upstream of most pre-MBT genes, there is H3K9/14ac in the pre-MBT embryos but not in the MBT and post-MBT embryos.

Together these results suggest that, unlike what has been found in mammalian cells and zebrafish embryos, both H3K4me3 and H3K9/14ac are highly associated with active transcription during MBT. They do not appear to be major pre-patterning markers for the massive ZGA. However, I found that there is a unique pattern of H3K9/14ac in pre-MBT embryos, whether it is maternally inherited and if it plays a role in instructing pre-MBT gene activation are open questions.

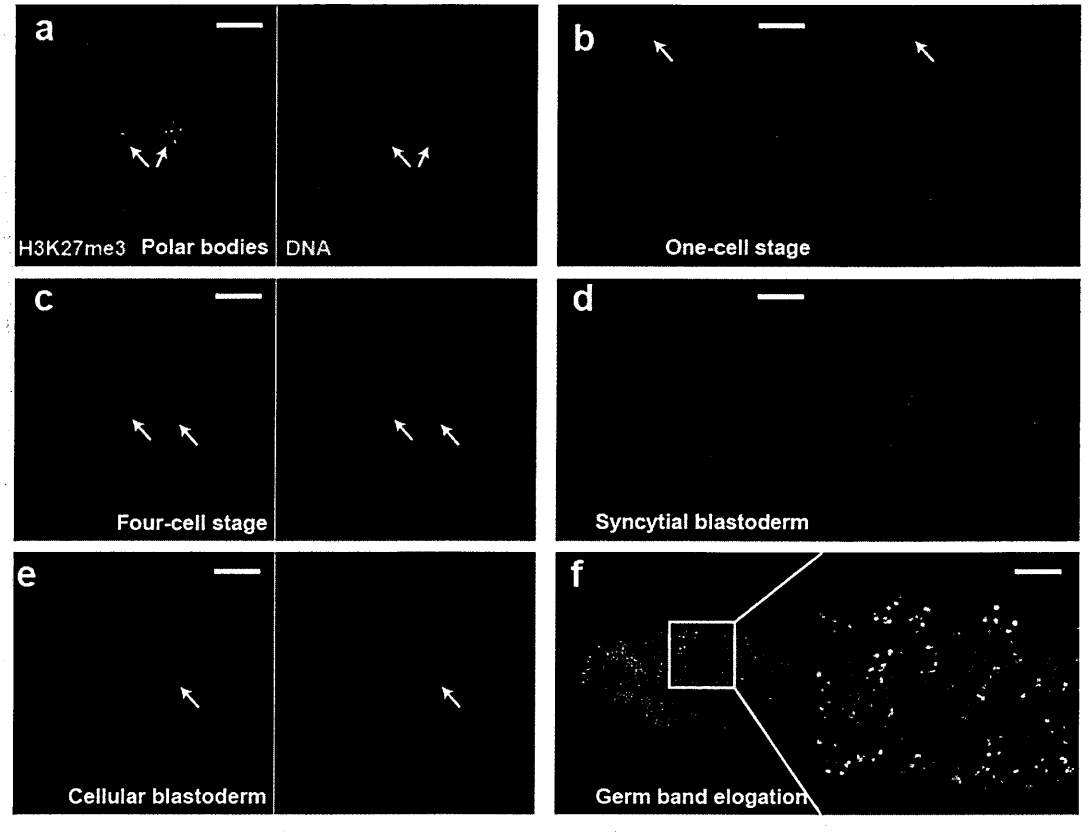


Figure 3-10. Re-appearance of H3K27me3 during *Drosophila* embryogenesis.
a,b,c, H3K27me3 can be detected in the nuclei and polar bodies (pb) of early pre-blastoderm nuclei but not during pre-MBT or in the somatic nuclei of MBT embryos **(d,e)**. H3K27me3 becomes detectable again in the germ cells of the late MBT embryos **(e)**, and becomes obvious after gastrulation **(f)**, in which a stripe pattern appears (scale = 20μm).

3.2.3 Re-establishment of H3K27me3 and Pc pattern during early embryogenesis

Next, I analyzed H3K27me3, another bivalent marker, which is deposited by PcG proteins, to test whether it might serve as an epigenetic marker to define the quiescent state

of developmental genes during *Drosophila* embryogenesis. As shown in Figure 3-10a,b,c, H3K27me3 can be detected in nuclei and in polar bodies at the earliest cleavage stages by immunostaining. This suggests that there is a maternal contribution of H3K27me3, which could potentially be maintained through the very early nuclear divisions.

However, H3K27me3 becomes undetectable in somatic nuclei in subsequent divisions and remains absent until after MBT (Figure 3-10d,e). At the end of cellularization, the H3K27me3 signal first reappears in the pole cells and accumulates until the pole cells invaginate. For the somatic nuclei, the reappearance of H3K27me3 starts after cellularization, which is consistent with recent findings (Petruk et al., 2012). The signal becomes apparent in some cells in a striped pattern during gastrulation. The H3K27me3 staining is more and more obvious as development progresses, and it expands to the majority of cells during segmentation (Figure 3-10f).

Consistent with the staining results, ChIP experiments for the H3K27me3 and Pc failed to recover any enrichment in the MBT embryos but successfully pulled down DNA at the PRE regions in the control post-MBT embryos (Figure 3-11). The global patterns of H3K27me3 and Pc are similar to previous ChIP-Chip/Seq studies, in which the H3K27me3 and Pc spread out in large regions defined as polycomb domains and have higher signals around or in the PRE, respectively (Schuettengruber et al., 2009). The average profile of H3K27me3 and Pc from different developmental time points also shows the gradually increasing spreading patterns around 488 known PREs over time.

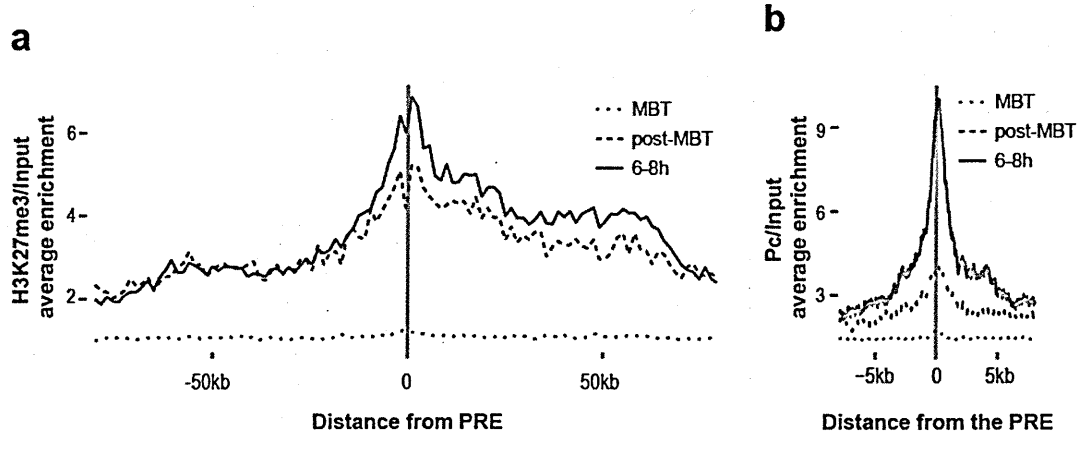


Figure 3-11. Re-establishment of H3K27me3 and Pc patterns during embryogenesis.
ChIP-Seq experiments confirm that **a,** H3K27me3 and **b,** Pc binding is absent during MBT and is gradually established after MBT. The average pattern of H3K27me3 and Pc surrounding 441 PREs is shown above (Polycomb Response Elements, see methods in Appendix).

Thus, H3K27me3 is likely present in oocytes but may be diluted or erased during replication. These dynamic changes in H3K27me3 and Pc during early embryogenesis are consistent with previous studies in that the PcG complexes start to function after cellularization (Denell, 1978; Lewis, 1978; Jürgens, 1985; Buchenau et al., 1998). These changes also argue against a role for H3K27me3 as paternal epigenetic marker, which was proposed recently in vertebrate studies (Hammoud et al., 2009).

3.2.4 Missing PcG-mediated long-range interactions in the early embryos

The absence of H3K27me3 and Pc binding in early embryos does not exclude the possibility that chromatin can retain other PcG components that maintain its quiescent state at developmental genes in the early embryos. Therefore, I performed chromatin FISH to test whether PcG-mediated long-range interactions exist in the early embryo.

As a control for the constitutive heterochromatin (Figure 3-12d,e,f), the pericentromeric repeats are away from the division plane during early divisions and cellularization (Dernburg et al., 1996). For probes against each locus of Hox clusters, ANT-C and BX-C, a transition from two dots per nucleus to single dot per nucleus is visible in pre-MBT and MBT embryos (Figure 3-12a,b). This observation agrees with previous studies that found that the onset of homologous pairing occurs during cellularization (Hiraoka et al., 1993; Fung et al., 1998; Gemkow et al., 1998), a signature for transvection (Pirrotta, 1990), which is assumed to play roles in promoting the transcription of homologous alleles.

However, in neither pre-MBT nor MBT embryos is there an obvious overlap of probes against two different Hox loci (Figure 3-12a,b), a marker for the chromatin kissing structure found in PcG proteins mediated silencing (Chambeyron, 2005; Kadauke and Blobel, 2009; Schuettengruber et al., 2009; Bantignies et al., 2011). This chromatin kissing is still obscure in the gastrulating embryos but becomes evident in some nuclei in the segmentation stage embryos (Figure 3-12c). Taken together, these results suggest that the PcG-mediated heterochromatin is established *de novo* after MBT, and matures during further differentiation. Higher nuclear organization, like the chromatin kissing structure, might be established even later in the nuclei with strong staining of H3K27me3.

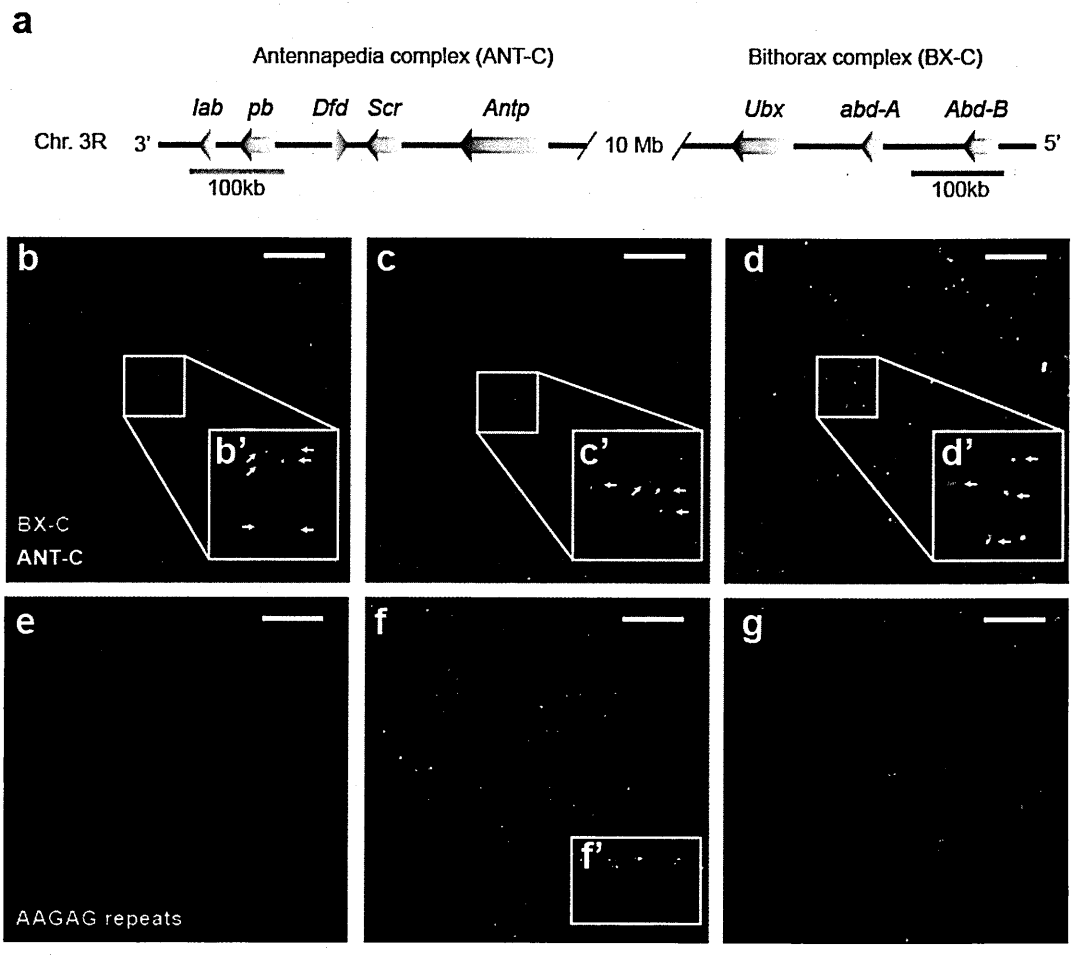


Figure 3-12. PcG mediated chromatin looping is established in post MBT embryos.
a, Schematic drawing illustrating the anterior and posterior Hox gene clusters in *Drosophila*. FISH probes for Antennapedia Complex (ANT-C, green) and Bithorax complex (BX-C, red) are represented at approximate locations. In **b** and **c**, colocalization of ANT-C and BX-C is very rare, and for each locus, there is a transition from two signals to one per nucleus. **d,** The overlap between two Hox clusters start to become obvious in some nuclei of the post-MBT embryos. **e, f,** and **g,** As pericentromeric repeats, the signals of AAGAG (red) are consistently detected at the apical side of the nuclei, which points away from the division plane (scale = 20μm).

3.3 High accessibility of chromatin in early embryos

3.3.1 Chromatin is loosely packaged in the early embryos

To understand how the chromatin is organized during early embryogenesis in the absence of PcG-mediated heterochromatin, I first performed micrococcal nuclease (MNase) gradient digestion for similar amounts of crosslinked chromatin isolated from pre-MBT, MBT, and post-MBT embryos (6-8 h) and analyzed the digestion pattern through agarose

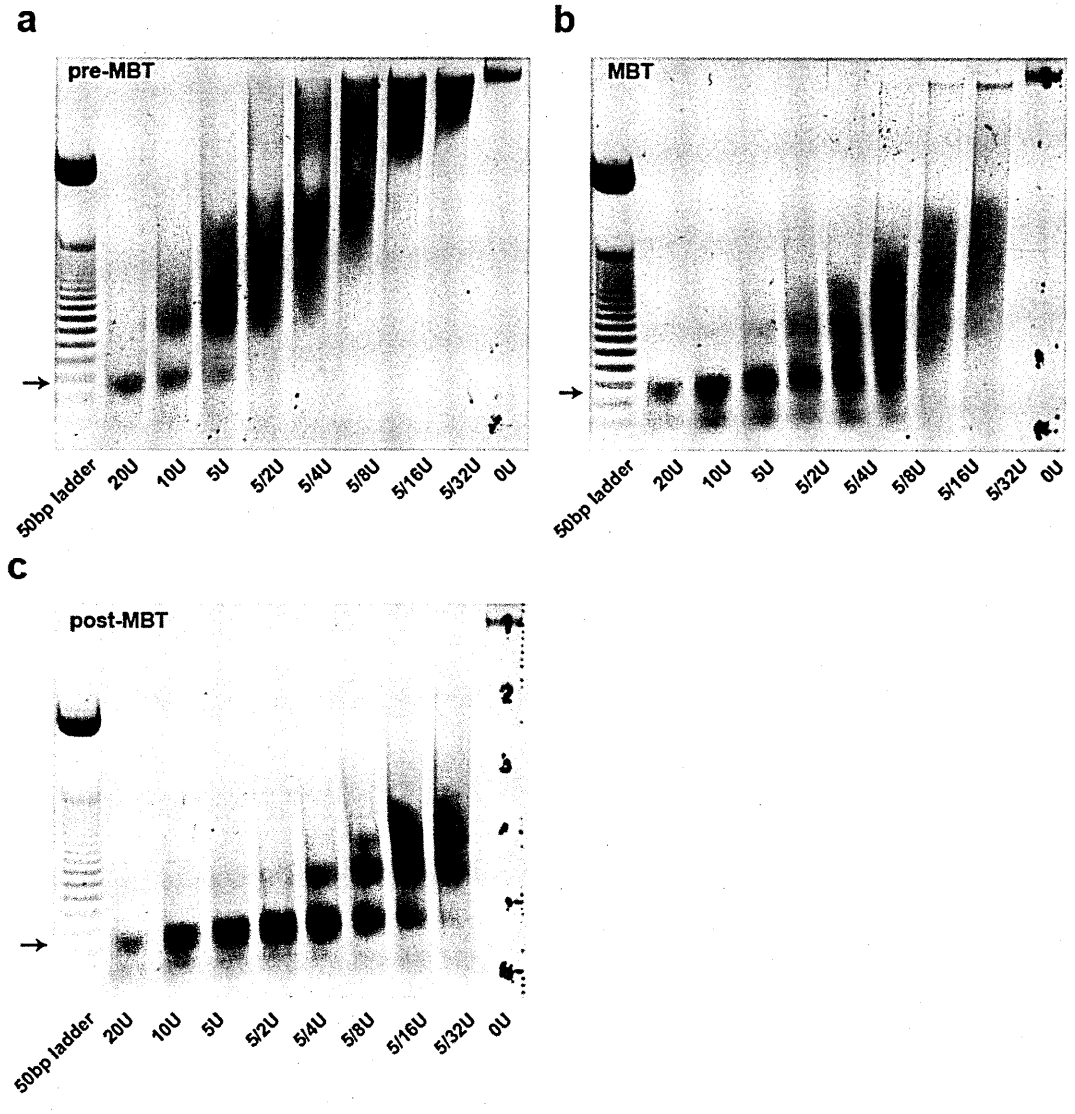


Figure 3-13. Different MNase digestion patterns of early and later embryos.
Nuclei of staged embryos were digested with gradient MNase and DNA was purified and analyzed on 1.7% agarose gel. **a,b,** mono- and di-nucleosome fragments, but not the multi-bands, can be obtained in the digested chromatin of early embryos (pre-MBT and MBT). **c,** As control, the ladder pattern of MNase digestion is clear in the later embryos (post-MBT, 6-8 h)

gel. The control (Figure 3-13c), a gradient MNase digestion for post-MBT chromatin, shows the typical ladder pattern as observed in previous studies (Spiker et al., 1983; Ross,

2001): from single-band (which represents mono-nucleosome) at highest concentration MNase to multi-bands (which represent mono-, di-, tri-, tetra-, and even penta-nucleosome) at the lowest concentration of MNase.

In contrast, the digestion patterns for pre-MBT and MBT chromatin are very fuzzy (Figure 3-13a,b) and the maximum bands obtained from gradient digestion are two for pre-MBT embryos and barely three for the MBT embryos. Furthermore, the digested patterns of pre-MBT and MBT samples look more resistant to MNase. One explanation is that embryos at different stages might be crosslinked at different levels: that is, early embryos could be easier to crosslink, and the over-crosslinked sample hinders the digestion. However, when I try to perform FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) to isolate regulatory elements, which are frequently nucleosome free (Giresi et al., 2007), a significant amount of DNA is obtained in the pre-MBT embryos. This suggests that the early embryos are not over-crosslinked since over-crosslinked chromatin would yield less free DNA that can be recovered through FAIRE. In addition, by performing a DNase I sensitivity assay in un-crosslinked embryos, Joey Nien in the Rushlow lab also found that the pre-MBT and MBT embryos are harder to digest than later embryos (personal communication). This also indicates that reasons other than crosslinking make the MNase digestion harder in the early embryos.

Another possible reason for the altered MNase pattern is that early embryo samples contain a large amount of yolk, which inhibits the MNase activity. In support of this, after purifying uncrosslinked nuclei by gradient centrifugation, early nuclei became more sensitive to DNase I, and the DNA gel results of digested pre-MBT embryos is smeared compare to those of post-MBT embryos (Lowenhaupt et al., 1983).

However, a more interesting hypothesis would be that the nucleosomes are simply not well positioned in early embryos, which may cause the distance between two nucleosomes to vary, leading to difficulties in positioning di- or tri- nucleosomes. In contrast to multiple nucleosomes, the chromatin can always be digested into mono-nucleosomes since there is always DNA wrapped around nucleosomes.

To further test the hypothesis that the early chromatin is relatively loose in the large nuclei of early embryos, I prepared samples from pre-MBT and post-MBT embryos for the transmission electron microscope. As shown in Figure 3-14a, in a pre-MBT embryo, the signal in the nucleus is evenly distributed without the dense structures that are frequently found at the periphery of normal nuclei and that are thought to be densely packed

heterochromatin. As a control, these dense structures can always be observed at the periphery and at the apical side of the nuclei in gastrulated embryos (Figure 3-14b). This indicates that heterochromatin is not yet fully formed in the early nuclei and supports the idea that chromatin is still loosely packed at this stage.

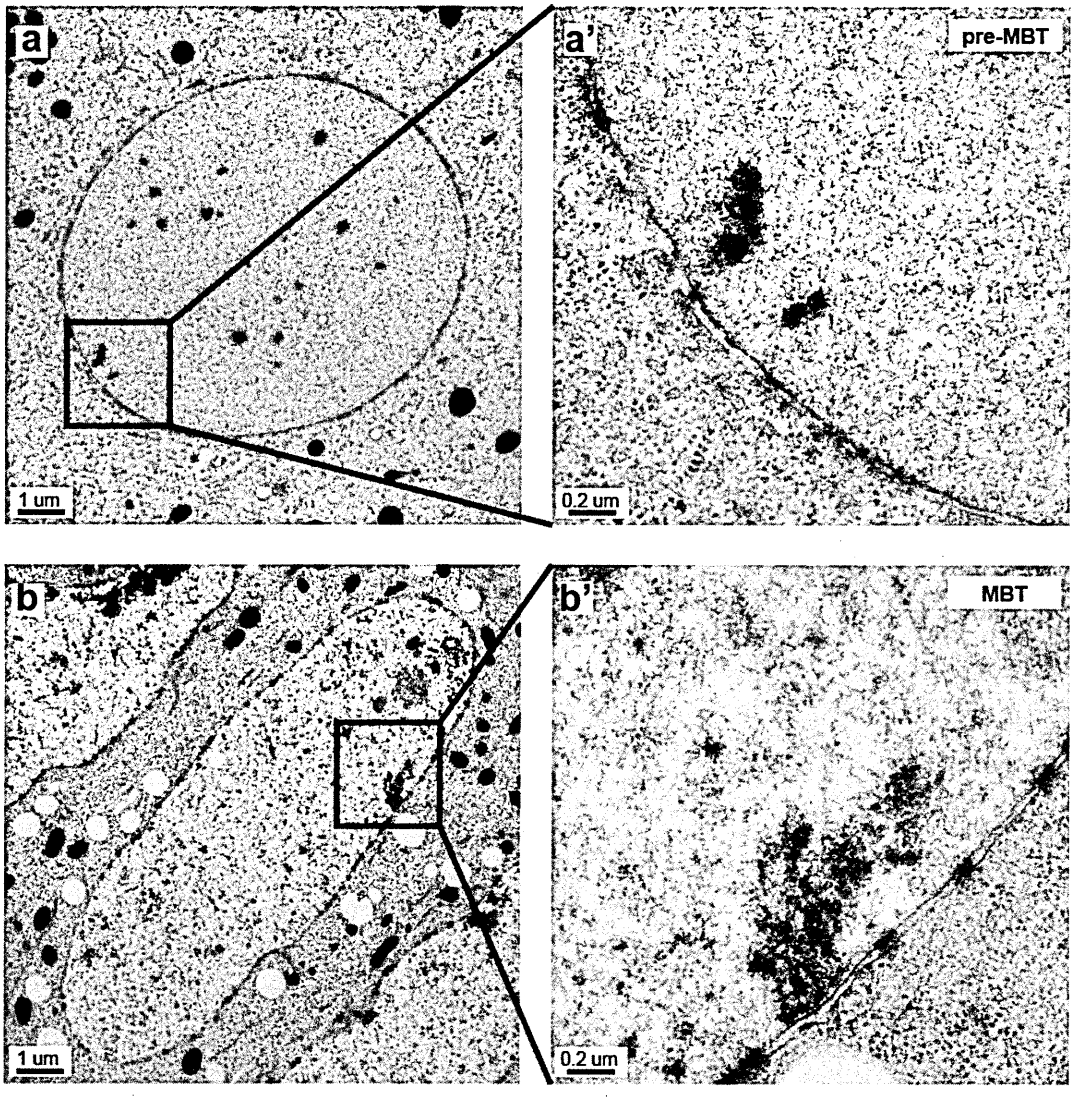


Figure 3-14. Small dense EM structure in nuclei of early embryos.
In **a** and **a'**, The density of the large round nucleus in the early embryo is very even. Despites a few very small, dense dots, at the periphery of the nucleus as **b** and **b'** in a post-MBT embryo there are no typical dense structures, which mark the heterochromatin.

3.3.2 Gradually defined nucleosome occupancy in the early embryos

To better understand how the nucleosome occupancy changes in early embryos, I prepared the MNase-digested samples for sequencing. For each time point, the lane with

only mono- and di-nucleosome bands was selected, and the mono-nucleosome band was cut out and prepared into a library for paired-end sequencing. Since the wrapping of DNA to nucleosomes is not as stringent as the binding of transcription factors, and since MNase is known to have sequence bias, the nucleosome occupancy reflected by MNase-Seq cannot be easily interpreted at the single-gene level in a consistent way (Yuan, 2005; Mavrich et al., 2008). Nevertheless, the average profile for a group of genes with certain criteria can be used to analyze their general trend of nucleosome occupancy. In this case, the pre-MBT genes (all), MBT active genes, MBT maternal genes (enriched for housekeeping genes), and silenced genes (constantly silenced during embryogenesis) were used for uncovering the changes in nucleosome occupancy during early embryogenesis.

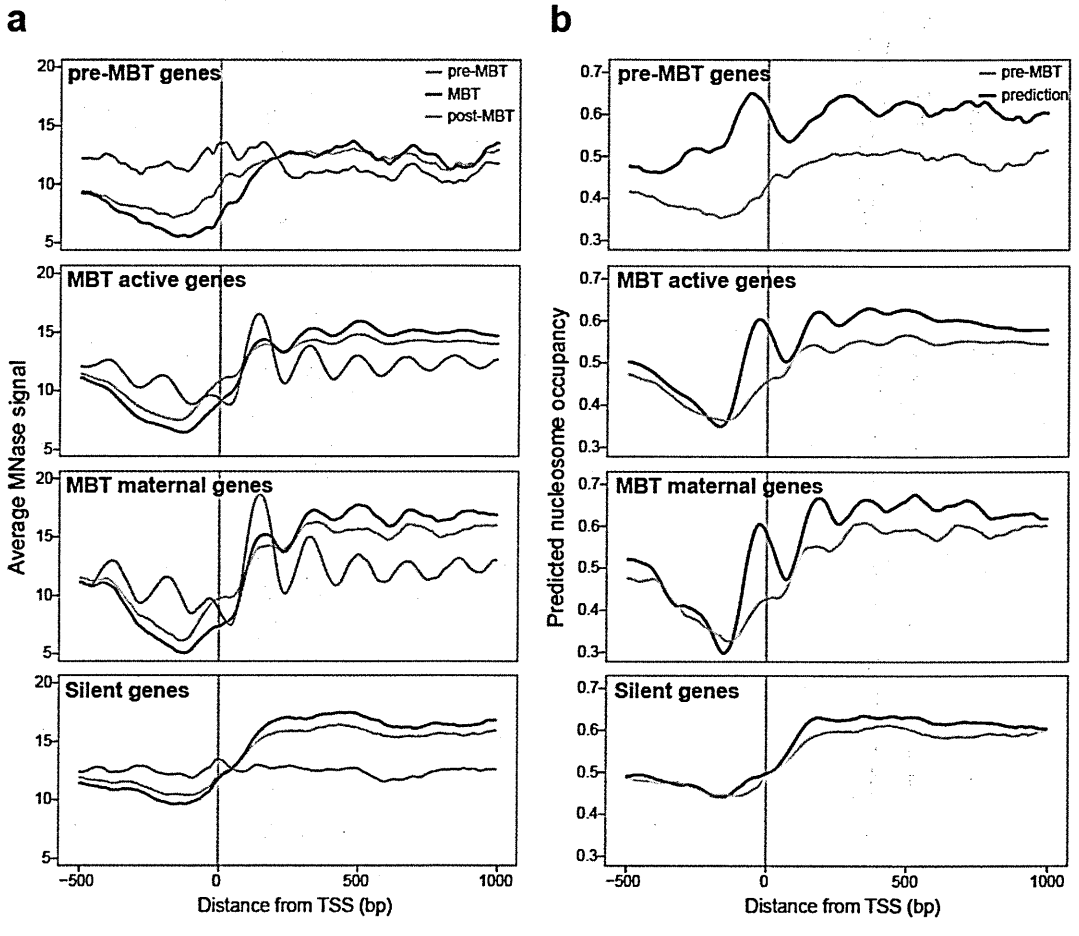


Figure 3-15. Nucleosome occupancy is gradually shaped during development.
a, Average distribution of nucleosomes around the TSS of different sets of genes displays the gradually phased pattern during development. **b,** The predicted nucleosome occupancy of different groups of genes based on the DNA sequence.

As shown in figure 3-15a, the nucleosome occupancy is generally fuzzy in all four groups in both pre-MBT and MBT samples, and it becomes more phased at MBT genes

(both active and maternal genes) at later developmental stages. In contrast, the nucleosome occupancy of pre-MBT genes and silenced genes stays fuzzy in the later embryos. This might be due to either small numbers of genes or the inactive states (Gilchrist et al., 2010).

Another interesting pattern for the early embryo is the reduced nucleosome occupancy at the promoter region in all four groups. For MBT genes, whose genes are transcribed at later time point, the nucleosomes increase in the promoter region during development and small nucleosome peaks become apparent. For pre-MBT genes and silenced genes, the nucleosome signal is high at the promoter regions at later stages, which correlates with the silent state of those genes.

Together with the DNA gel findings, the MNase-Seq results support the idea that the early chromatin is not well phased by nucleosomes and might be highly accessible for developmental regulators.

3.3.3 Nucleosome occupancy in the early embryo is determined by both DNA sequence and transcription factor binding

To further evaluate the effect of DNA sequence on nucleosome organization during early embryogenesis, we compared the predicted nucleosome pattern around the promoter region of the different gene groups. These predictions are made from sequence alone and are based on an algorithm developed by John Widom's lab (Kaplan et al., 2009).

Surprisingly, the MNase-digested pattern in the early embryos is generally very similar to the predictions, especially for silenced genes (Figure 3-15b). The only significant difference for the activated groups (pre-MBT and MBT genes) is the depletion of the predicted –1 nucleosome in the early digested profiles. This depletion is consistent with previous findings and might be due to the presence of Pol II and other transcription factors (Mito et al., 2005; Schones et al., 2008; Gilchrist et al., 2010). However, this similarity decreases in later embryos as the nucleosome occupancy becomes more localized for the activated genes, and the nucleosomes fill the promoter regions of silenced genes. This indicated that the initial nucleosome position is determined by both DNA sequence and binding of transcription factors at the early beginning. However, the stable transcription of the activated genes and higher chromatin architecture formed in later stages play roles in confining the nucleosome occupancy and chromatin accessibility during development.

3.4 Different usage of core promoter motifs during ZGA

3.4.1 Massive Pol II recruitment during MBT is not dependent on a few early transcription factors

What determines the differential activations of pre-MBT and MBT genes during ZGA if there is little chromatin influence? One explanation is that some early developmental regulators that are deposited maternally trigger the transcriptional cascade that carries out the massive ZGA during cellularization. In this case, the recruitment of Pol II to MBT genes would be affected if some early factors were absent.

To test this hypothesis, Nina Keonecke (another graduate in our lab) and I first performed Pol II ChIP-Seq on gd7 mutant embryos, which lack Dl activity due to the inactivation of the Toll signaling pathway (Konrad et al., 1988). By comparing the Pol II pattern of the Dl target genes in the gd7 mutant to that of wildtype, I found that, on average, there is less Pol II at the transcription unit in the gd7 mutant than in wildtype, which suggests that Dl may play role in releasing paused Pol II for elongation. However, there is no significant decrease in Pol II recruitment in the MBT embryos at the TSS in gd7 compared to wildtype (Figure 3-16a).

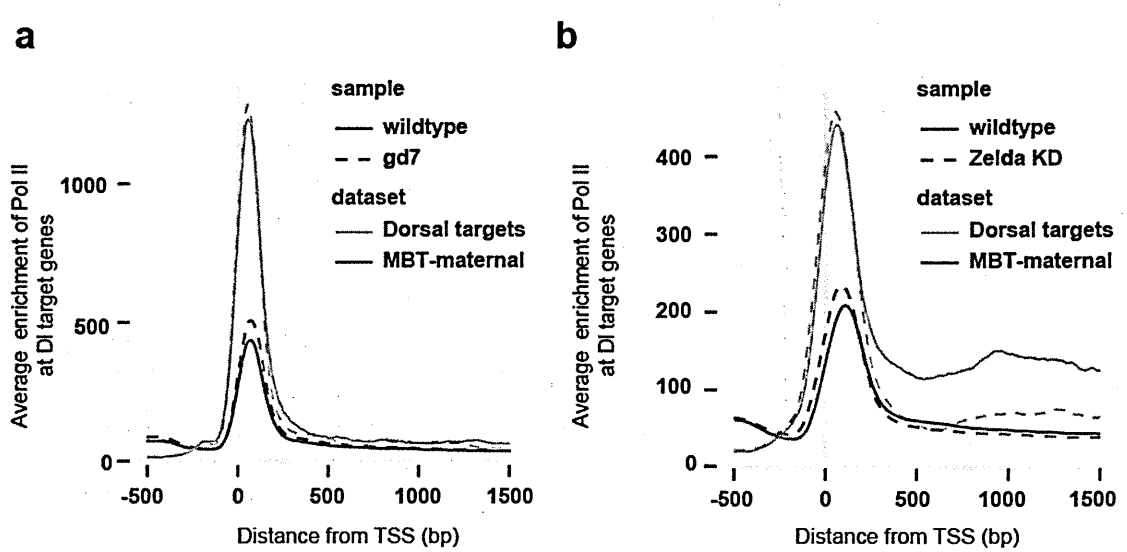


Figure 3-16. Global Pol II recruitment to Dl targets is not affected in Dl or Zld depletion. Average Pol II ChIP-Seq pattern at Dl targets shows the decreasing signal of Pol II along the gene body with the depletion of Dl (Red dash line in **a**) or Zld (Red dash line in **b**). However, compared with wildtype control, the recruitment of Pol II to the TSS is not compromised.

A recent study revealed that Zelda (Zld), a maternally deposited zinc-finger protein, is critical for activating the *Drosophila* genome by binding a cis-regulatory heptamer motif, CAGGTAG, shared in many early zygotic genes (Bosch et al., 2006; Liang et al., 2008). In collaboration with Lydia Sun from the Rushlow lab, I further examined Pol II recruitment of Dl targets in Zld knockdown MBT embryos. Although some early genes involved in neurogenesis cannot recruit Pol II, most Dl target genes affected in Zld depleted embryos, similar to Dl depletion, failed to elongate and showed Pol II pausing (Figure 3-16b). This suggests that maternally deposited developmental regulators, at least for Dl and Zld, are not the key for the differential Pol II recruitment during ZGA. Although it is possible that other transcription factors play redundant roles in recruiting Pol II to those Dl target genes, this hypothesis cannot explain why most housekeeping genes recruit Pol II during MBT but not pre-MBT.

3.4.2 Different usage of core promoter motifs during ZGA

Although there is no bivalent domain for instructing the differential ZGAs, the generally low level of histone modifications in pre-MBT genes suggests that there might be some general regulatory difference between pre-MBT and MBT genes. A recent genomic study in mammalian cells found that genes with focused promoters have less H3K4me3 than genes with dispersed promoters (Rach et al., 2011). This raises the possibility that different core promoters might be used for recruiting different general transcription factories to activate pre-MBT and MBT genes with different chromatin signatures.

To test this hypothesis, Jeff and I compared our gene groups with predicted core promoter motifs (Ohler et al., 2002; FitzGerald et al., 2006; Hendrix et al., 2008). Consistent with a previous study (Ni et al., 2010), MBT-maternal genes are strongly enriched for core promoter elements associated with dispersed initiation such as Ohler1, 6, 7, and DRE (Dref response element). In contrast, highly regulated genes, like pre-MBT genes and MBT zygotic genes, are highly represented with core promoter elements associated with focused initiation.

Inr, DPE, MTE and PB (Pause Button) have previously been associated with paused genes (Ohler et al., 2002; Hendrix et al., 2008). Indeed, MBT active genes and MBT paused genes are highly enriched for these elements. Both groups are also enriched for GAGA, consistent with reports that GAF promotes the recruitment of paused Pol II

(Lee et al., 1992; Leibovitch et al., 2002; Lee et al., 2008). We noticed that these genes are not significantly enriched for the TATA box, although they are usually occupied by TBP (Figure 3-18a, and average pattern in Figure 3-6a).

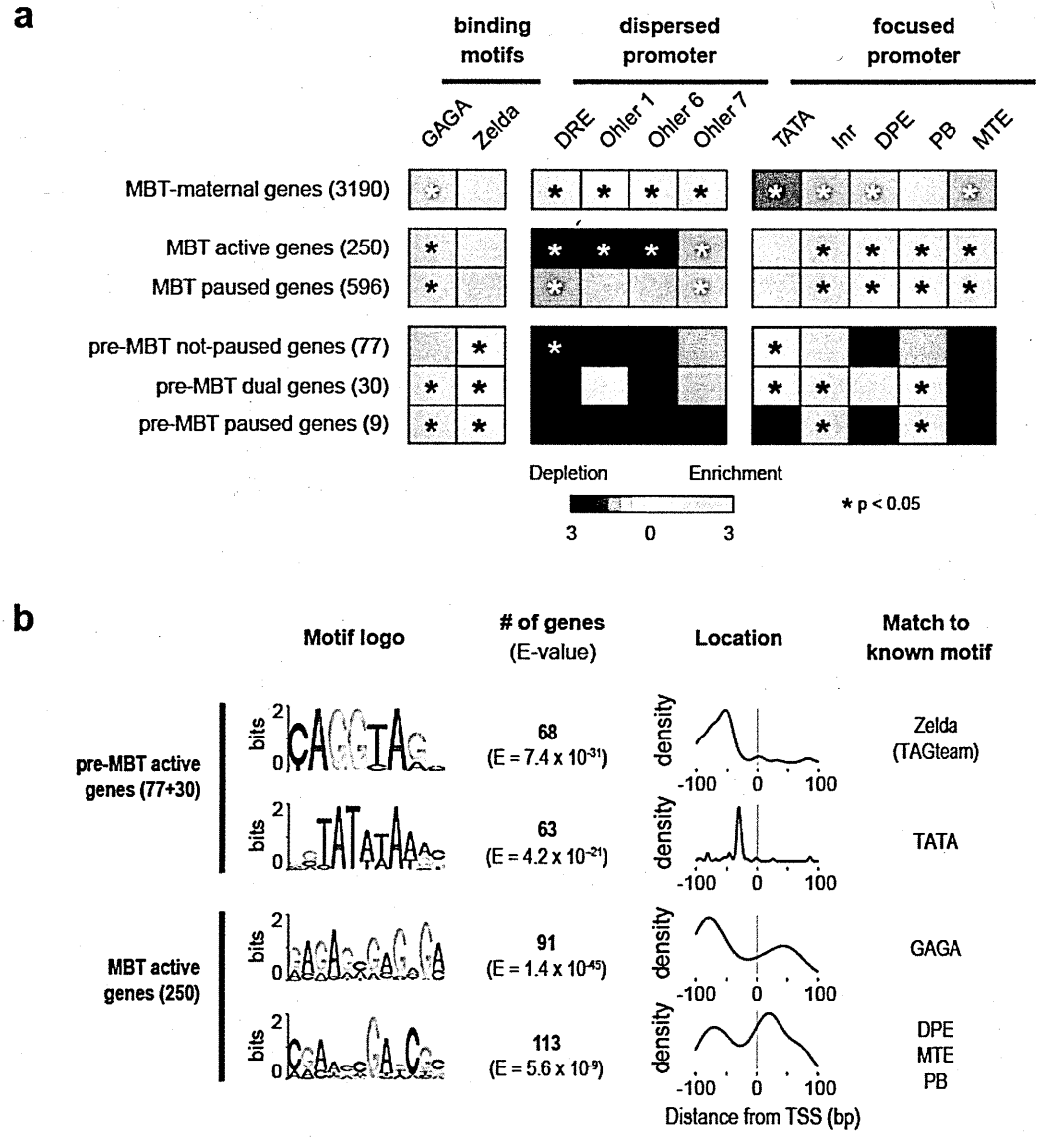


Figure 3-17. Differential usage of core promoter elements during the zygotic genome activation.
a, Analysis of known core promoter elements and binding motifs for GAGA factor (GAF) and Zelda. The asterisk indicates that the enrichment (yellow) or depletion (black) is significant ($p < 0.05$).
b, The top two known motifs identified by *de novo* analysis (MEME) in pre-MBT and MBT genes, with their locations relative to the TSS (shown as density distributions).

TATA box is significantly enriched in promoters of pre-MBT genes (Figure 3-17a). However, only pre-MBT genes that initially show the non-paused profile (pre-MBT not-paused, pre-MBT dual) are significantly enriched for Inr and TATA, while the pre-MBT paused group is not enriched for TATA. Furthermore, only pre-MBT genes that are paused at some point (pre-MBT dual, pre-MBT paused) show enrichment for GAGA, Inr and PB,

while the pre-MBT not-paused group is under-enriched for PB. Enrichment of Zld motifs, on the other hand, is found in all groups of pre-MBT genes. Thus, during pre-MBT stages, the presence of Zld motifs correlates with the recruitment of Pol II to the promoter, while the presence of specific core promoter elements correlates with the Pol II occupancy profile across the gene body.

To further consolidate the differences between pre-MBT and MBT genes, we performed *de novo* motif analysis with MEME on the 200 bp centered on the transcription start site (Figure 3-17b). For pre-MBT stage non-paused genes (pre-MBT not-paused, pre-MBT dual), the top two motifs were Zelda and TATA. Thus, Zld motifs are enriched very close to the transcription start site (~50 bp upstream, Figure 3-17b). In contrast, the top two known motifs for the most comparable MBT group (MBT active genes, which are also early-expressed developmental genes) were GAGA and a motif that resembles DPE, MTE and PB.

This suggests a model in which Pol II is recruited to pre-MBT genes through TATA box in the promoter, and rapid transcription without Pol II pausing is mediated through a TATA-enriched promoter (Figure 3-18). In contrast, paused Pol II is typically established through GAF during MBT at promoters with pausing elements such as DPE, MTE or PB. Thus, there are two principle modes of Pol II recruitment and transcription, but genes may also have elements from both and show a dual behavior.

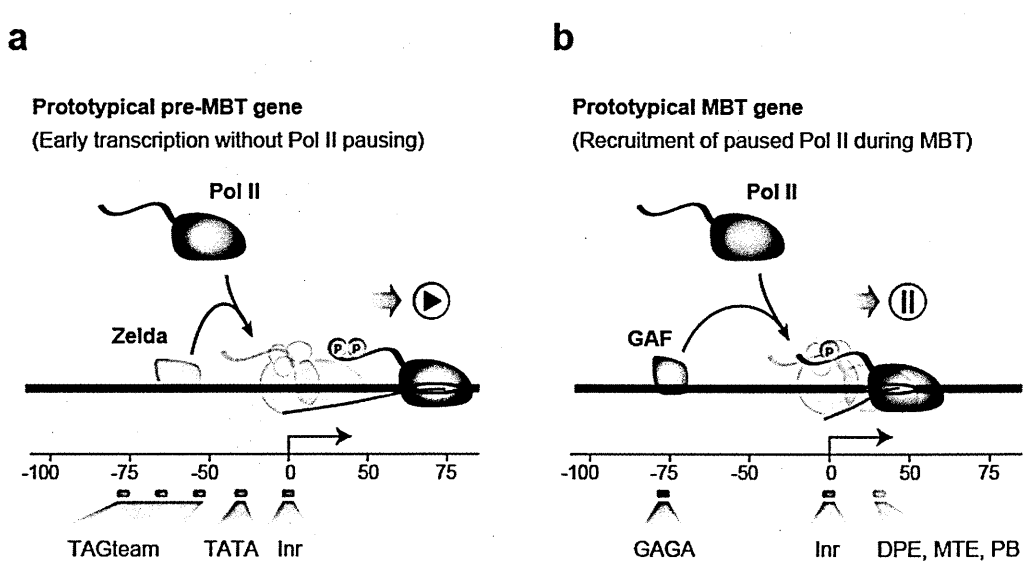


Figure 3-18. Model for Pol II recruitment during ZGA.
Proposed model for Pol II recruitment and elongation behavior at prototypical pre-MBT (**a**) and MBT (**b**) genes during the zygotic genome activation. Genes can have elements of both modes. Transcription factors in addition to Zelda and GAF may also influence the Pol II behavior at genes.

CHAPTER 4

Discussion

4.1 Discussion and Speculations

4.1.1 Transcriptional strategies in *Drosophila* ZGA

In this study, I successfully established a ChIP-Seq system for early embryos that uses a reasonable amount of sample material (2,000,000 *Drosophila* nuclei). Compared with previous genome-wide RNA studies (De Renzis et al., 2007; Lott et al., 2011), these ChIP-Seq results more precisely reflect the zygotic transcription of the embryonic nuclei without maternal RNA influence. It not only can identify which genes are transcribed during embryogenesis, but it can also distinguish exactly which promoters are used for genes with multiple promoters (Figure 3-4). Based on the Pol II and TBP ChIP-Seq results, I uncovered two transcriptional strategies used during *Drosophila* ZGA: non-paused transcription for ~110 pre-MBT genes in the early divisions and widespread Pol II pausing for MBT genes established during global ZGA. The significant difference in Pol II pausing along with different core promoter usages between pre-MBT and MBT genes likely serves different purposes during early *Drosophila* development.

Why are genes non-paused during pre-MBT divisions?

Pre-MBT genes that are required for very early developmental events are presumably optimized to achieve high levels of transcripts during the extremely short mitotic cycles before MBT (Foe et al., 1993) and the abortion of nascent transcripts caused by quick mitosis (Shermoen and O'Farrell, 1991; Rothe et al., 1992). In addition to the particularly short genes, the TATA-containing promoter may also be critical for fast transcription. TATA is a strong core promoter element that efficiently supports transcription *in vitro* (Aso et al., 1994), mediates efficient re-initiation *in vitro* (Yean et al., 1997; 1999), and its presence *in vivo* correlates with 'bursts' of transcription that produce many transcripts within a short time (Zenklusen et al., 2008). Furthermore, it has been shown that TATA promotes pTEFb activity, leading to more efficient elongation rates *in vitro* and *in vivo* (Amir-Zilberstein et al., 2007; Montanuy et al., 2008). A possible disadvantage of TATA- mediated transcription is that it may be more stochastic (Hornung et al., 2012). However, this is unlikely to be harmful at pre-MBT stages since the *Drosophila* embryo is still a syncytium, and gene products have a chance to diffuse.

In contrast, although Pol II pausing may have advantages for precise and synchronous expression in response to localized extracellular signals (Boettiger and Levine, 2009), it could be specifically unfavorable for the pre-MBT stage embryos. A

recent study on *Drosophila* follicle cells found that the DNA replication at certain loci is compromised when a transcriptional inhibitor (α-amanitin) was added to block transcription elongation (Xie and Orr-Weaver, 2008). This replication defect may be due to the collision between the replication complex and the immobilized Pol II on the DNA (Gu et al., 1993). In this case, the absence of Pol II pausing in the pre-MBT embryos is advantageous for preventing the genomic instability caused by transcription and DNA replication collision during the early quick divisions (Kim and Jinks-Robertson, 2012).

The strategies of using different core promoters with different Pol II patterns might be conserved in the other species with rapid pre-MBT division. Indeed, functional studies of TBP in both zebrafish and *Xenopus* embryos have shown that, rather than being essential for global ZGA, TBP is required for activating a subset of genes (Prioleau et al., 1994; Müller et al., 2001). This indicates that TATA-box, or some other strong motif directly recognized by TBP, can be functionally conserved for pre-MBT gene activation, although this assertion needs further study to confirm.

Can paused Pol II reflect transcriptional potential?

In this study, I find that widespread Pol II pausing, which has been previously found in *Drosophila* peri-gastrulation embryos (Zeitlinger et al., 2007), is *de novo* established during MBT. Among the paused genes, a large portion of quiescent developmental genes will be activated at certain timepoints after MBT. This finding is similar to recent observations in zebrafish that the pluripotent chromatin state, bivalent domains and paused Pol II at developmental genes, is established during MBT (Vastenhouw et al., 2010). In addition, in mammalian ESCs, quiescent developmental genes with short nascent RNAs, a marker of Pol II pausing, and bivalent domains can be expressed during differentiation (Guenther et al., 2007). Similar to the dynamic changes in bivalent chromatin during differentiation observed in mammalian cells (Vastenhouw and Schier, 2012), Bjoern Gaertner, another graduate student in our lab, showed that the recruitment of Pol II pausing is stage-specific and highly associated with transcription potential during myogenesis (Gaertner et al., 2012).

These similarities imply that widespread paused Pol II at quiescent developmental genes may reflect the transcription potential in responding to extracellular signals and associated with the developmental competence of the early MBT cells. Interestingly, in *C. elegans*, which does not have the NELF complex or classical Pol II pausing, , the cellular

lineage is well defined during development. Furthermore, compared with other systems, natural trans-differentiation, de-differentiation, and stem cells are very rarely found in *C. elegans* (Hajduskova et al., 2011; Tursun, 2012). In contrast, the planarian, another primitive triploblastic model organism, has strong regeneration ability and possesses homologs for most subunits of the NELF complex [(Alvarado, 2003), communicated with Longhua Guo in Sánchez Alvarado lab].

4.1.2 *de novo* establishment of chromatin landscapes during MZT

My immunostaining and ChIP-Seq results clearly suggest that there is no bivalent domain for instructing ZGA and most chromatin signatures are established *de novo* during early *Drosophila* development. These dynamic patterns of some histone modifications are very similar to recent observation in *Xenopus* (Akkers et al., 2009), which have shown that H3K27me3 is gradually deposited after gastrulation. This indicates that at least *Drosophila* and *Xenopus* may not follow the model recently proposed in vertebrate sperm studies, in which bivalent domains in sperm may instruct early development. It is possible that the pluripotent chromatin state or pre-patterning markers might exist in different forms due to species specificity. However, it is also possible that the basic mechanisms underlying the establishment of chromatin landscapes may be evolutionary conserved in vertebrates, or even across all metazoans. Thus, histone modifications found in sperm may not be instructive for ZGA.

H3K4me3, the lost memory in invertebrates?

In agreement with a recent work on epigenetic maintenance in *Drosophila* embryogenesis (Petruk et al., 2012), I do not observe H3K4me3 in early *Drosophila* embryos until the end of MBT. This suggests that H3K4me3 cannot serve as an epigenetic marker for early gene activation. Moreover, even in MBT and post-MBT embryos, H3K4me3, like H3K9/14ac, is only found further downstream of the TSS of MBT genes that are actively transcribed. Together with our recent observation in myogenesis that H3K4me3 is only associated with active genes during time courses (Gaertner et al., 2012), H3K4me3 more likely reflects the active state, rather than predicting transcription potential, of a gene in *Drosophila.*

However, H3K4me3 has been found at the TSS of inactive genes, with or without Pol II, in many mammalian studies, and most of those genes can be activated during differentiation (Bernstein et al., 2006; Guenther et al., 2007; Mikkelsen et al., 2007; Pan et

al., 2007; Cui et al., 2009; Lien et al., 2011). In addition to its genomic association with transcription potential in vertebrate systems, biochemical studies demonstrate that H3K4me3 can be recognized by TAF3 directly via its PHD domain (Vermeulen et al., 2007; van Ingen et al., 2008). This provides the molecular basis for the hypothesis that H3K4me3 may serve as an epigenetic marker for instructing gene activation. Although TAF3 is highly conserved in eukaryotes, the sequences of its PHD fingers in invertebrates are very different from the homologues in vertebrates and other PHD-finger-containing proteins (van Ingen et al., 2008; ensembl). This difference may be the reason why H3K4me3 could be an epigenetic marker for vertebrates but not for *Drosophila*. By examining the relationships among H3K4me3, TAF3, and Pol II in animals of other phyla, we might work out a better understanding of how the H3K4me3 evolved as an epigenetic marker in metazoans: either newly evolved in vertebrates or gradually deteriorated in invertebrates.

Re-establishment of PcG mediated silencing, a conserved event in metazoan development

Although H3K27me3 appears to be inherited by maternal deposition in *Drosophila* early nuclei, the diminishing signal during subsequent divisions suggests that H3K27me3 may just reflect the function of PcG during gametogenesis, rather than serve as an epigenetic marker for instructing the formation of PcG-mediated silencing. Unlike previous findings in *Drosophila* imaginal discs, which find that PcG proteins, including Pc, can always bind to the PRE site despite the transcriptional state (Papp and Müller, 2006), my ChIP-Seq and chromatin FISH results suggest that the binding of Pc and the looping between two HOX loci are concurrent with H3K27me3 deposition after cellularization. This is consistent with a recent study, in which the formation of polycomb bodies are found to start at the end of cellularization and become more stable during later development (Cheutin and Cavalli, 2012). These results are also in agreement with a previous study showing that the PcG mutants mostly affect the body patterning after gastrulation (Denell, 1978; Lewis, 1978; Jürgens, 1985; Buchenau et al., 1998). This also explains why the depletion of maternal PcG components only increases the AP patterning transformation in zygotic mutants, but cause minimal effect and can be rescued with zygotic proteins (Haynie, 1983). The mitotic defects observed in some PcG mutants might be caused by secondary effects of the maternal depletion of PcG during oogenesis (Phillips and Shearn, 1990; O'Dor et al., 2006).

Recent genomic studies on mammalian sperm and ESCs found overlapping of H3K27me3 at developmental genes in sperm and ESCs, which lead to the proposal of an intriguing model in which H3K27me3 might serve as an epigenetic signature that can be transmitted from one generation to another. The limits posed by embryo availability and the number of nuclei per embryo makes this hypothesis extremely hard to test by ChIP-Seq, and it still leaves a knowledge gap between gametes and ESCs. However, through immunostaining, researchers did find significant decreases of H3K27me3 in mouse morula stage embryos and re-deposited of H3K27me3 to cells of the inner cell mass at the blastocyst stage (Bao et al., 2005; Zhang et al., 2009a). Rather than passive dilution through division, the dramatic change indicates that it is highly regulated by histone demethylation. Although the immunostaining results cannot tell us exactly which loci lose the H3K27me3 marker during mammalian early development, it does suggest that a large portion of the chromatin re-establishes the PcG mediated silencing. This is consistent with recent ChIP-Chip study in *Xenopus* early embryos (Akkers et al., 2009), which were unable to detect the H3K27me3 pattern until gastrulation. All of this evidence together suggests that the re-establishment of PcG-mediated silencing is a conserved developmental event from invertebrates to vertebrates, and the overlap of the H3K27me3 between sperm and ESCs may just be a correlation.

H3K9/14ac, a connection between replication and pre-MBT gene activation?

Remarkably, unlike H3K4me3 and H3K27me3, the H3K9/14ac is continuously detectible early on with a distinct pattern in pre-MBT embryos, that is, H3K9/14ac is deposited a few thousands base-pairs upstream of pre-MBT genes rather than around the TSS of activated genes in later embryos. Although it is unlikely that H3K9/14ac can be a chromatin landscape for predicting MBT gene activation, due to the technical difficulty, it is hard to test whether this is an inherited epigenetic marker for instructing pre-MBT gene activation.

How is H3K9/14ac maintained in the early mitosis without transcription and what is the role of H3K9/14ac in the early embryos? A recent study on Gcn5, the enzyme that mediates H3K9/14ac, suggests that it is required for efficient deposition of new H3 with the H3K9/14ac onto replicating DNA (Burgess et al., 2010). This DNA replication dependent H3K9/14ac model explains why the H3K9/14ac can be maintained during a quick division without any transcription. Moreover, it is also consistent with our

observation that H3K9/14ac upstream of the promoter region diminishes in the post-MBT embryo when nucleic division slows down.

Is there any connection between DNA replication and pre-MBT gene expression in the early embryo? Although there is no evidence to show whether the replication origins can regulate transcription, genomic studies of the replication origins found that there is a strong correlation between the origin recognition complex (ORC) binding sites and active promoters in the open chromatin (Aggarwal and Calvi, 2004; MacAlpine et al., 2004; 2010). Despite the assumption that origins of DNA replication could general start everywhere on genome during the early division, the H3K9/14ac activity of ORC2 complex, and the overlap between H3K9/14ac and pre-MBT genes raises the possibility that the replication origins might play roles in the transcription of pre-MBT genes. For example, the H3K9/14ac upstream of the promoter region might make the chromatin more accessible to transcription factors and facilitate gene activation. Interestingly, a recent genomic study in zebrafish found that H3K14ac in sperm is highly associated with pre-MBT genes (Wu et al., 2011), which implies that the link between H3K9/14ac and pre-MBT gene activation may be conserved in other systems.

4.1.3 Nucleosome organization during early embryogenesis

DNA gel and genomic sequencing results followed by gradient MNase digestion imply that the early chromatin is not well packaged by nucleosomes. This loosely-packed chromatin in the early embryos may provide a great system to understand how nucleosome positioning is determined during development and before or after transcription.

Determinants of nucleosome position

Intriguingly, the fuzzy nucleosome occupancy pattern obtained through MNase-Seq for the early embryos closely resembles the nucleosome occupancy pattern predicted based on DNA sequence. This observation is very similar to recent findings in yeast (Kaplan et al., 2009; Zhang et al., 2009b; 2011), in which the nucleosome occupancy of the *in vitro* assembled chromatin is consistent with the predicted model. These results indicate that the intrinsic histone-DNA and transcription factor-DNA interactions are the major determinant of nucleosome positions at the early beginning, when the chromatin is loosely packaged without higher chromatin architecture. During cell differentiation, there are more and more factors, such as chromatin remodeling complexes, insulators, and PcG complexes, involved in refining the nucleosome organization and forming the complicated

sub-nuclear compartments. This makes the intrinsic DNA sequence less effective at predicting the nucleosome positions *in vivo* (Zhang et al., 2009b).

Nucleosome position and paused Pol II

By comparing the nucleosome occupancy and Pol II binding profile of *Drosophila* embryos, Mavrich et al. observed an overlap of paused Pol II with the +1 nucleosome and speculated that the +1 nucleosome might cause Pol II pausing by imposing a stable barrier to elongation (Mavrich et al., 2008). In contrast, a recent study on *Drosophila* S2 cells found that genes with higher levels of paused Pol II have less +1 nucleosome occupancy and argue that, rather than the +1 nucleosome causing the paused Pol II, the paused Pol II might disrupt DNA-specified nucleosome organization and play a role in maintaining the euchromatin state for future gene activation (Gilchrist et al., 2010). The latter is consistent with my observation that the nucleosome occupancy pattern in MBT embryos is very fuzzy and similar to the prediction, although widespread Pol II pausing is already established. This suggests that the +1 nucleosome cannot be the leading reason for the poising of Pol II.

After further development, the nucleosome occupancy is shaped at genes that recruited Pol II during MBT. How is this achieved? Recent *in vitro* yeast chromatin reconstitution experiments provide an explanation (Zhang et al., 2011). The chromatin reconstituted *in vitro* with purified histones cannot recapture the nucleosome occupancy pattern *in vivo*. However, the pattern can be shaped as it is *in vivo* after the addition of yeast nuclear extract and ATP, but not the yeast nucleic extract only. This suggests that ATP-dependent remodeling factors, which are involved in transcription, might play important roles in the nucleosome shaping during development.

4.1.4 A general model for the timing of MZT

Although several mutually compatible molecular mechanisms have been proposed to explain the regulation of MZT (Tadros and Lipshitz, 2009), most of them ignore the differential activation of pre-MBT genes and MBT genes. The results of this genomic study not only give us a better understanding of the role of chromatin in transcription regulation during early *Drosophila* development, but also provide me an opportunity to propose a general model to explain the differential onset of ZGA (Figure 4-1.).

The onset of pre-MBT genes: beyond early activators

In early *Drosophila* embryos, the initial chromatin is loosely packed and highly accessible to transcription factors. However, in the pre-blastoderm embryo, TBP is maternally deposited as mRNA and can be detected by immunostaining until the beginning of pre-MBT. This suggests that the availability of TBP protein could be the key reason for the beginning of transcription in the pre-MBT embryo. Together with early activators (*e.g.* Zld and Dl), TATA-box containing promoters of pre-MBT genes may recruit limited TBP with higher affinity than TATA-less promoters in the pre-MBT embryos. In addition, H3K9/14ac deposited during DNA replication may also boost early transcription by recruiting chromatin remodeling factors through their bromodomain. These together give the pre-MBT genes the opportunity to be transcribed earlier without Pol II pausing.

miRNA, an additional link between pre-MBT genes and MBT genes

Among ~110 pre-MBT active genes, there are two miRNA clusters, one of which (*mir-309*) has been previously reported to be an early zygotic gene that mediates maternal RNA degradation in a SMG-dependent manner (Bushati et al., 2008; Benoit et al., 2009). It is consistent with the previous finding that the massive maternal degradation occurs approximately 2h after fertilization, following the onset of zygotic transcription (Bashirullah et al., 1999). The gradual degradation of maternal mRNA coding repressors, like Tramtrack [TTK, a known repressor that regulates ftz in a N/C ratio dependent manner (Pritchard and Schubiger, 1996)], may function as a maternal clock to release the silencing of most MBT-zygotic genes, the majority of which occur by GAF dependent activation by freeing their targets to GAF during MBT (Pagans et al., 2002; Lu et al., 2009).

The missing general factor, another cause of global ZGA

The absence of Pol II recruitment at most housekeeping genes in pre-MBT embryos indicates that there are different usages of general transcription factors during ZGA. Indeed, certain mutants of Pol II subunits (*e.g. wimp*), and even TBP, only affect small portions of early developmental genes but not housekeeping genes [(Parkhurst and Ish-Horowicz, 1991), and unpublished data from the Ming-Tsan Su lab]. This implies the existence of additional general factors, which are not translated or activated until MBT, for activating housekeeping genes with weak promoters.

For pre-MBT genes with TATA-box-containing promoters, TBP can be directly recruited and further assembles the PIC at the promoter for transcription. In contrast, for

those genes activated later whose promoters do not contain a TATA-box, TBP cannot be directly recruited, and these genes must therefore adopt different ways to assemble the PIC with additional GTFs. For this reason, these genes may not be expressed until MBT. Besides the core promoters, which I believe have general roles in timing the activation of pre-MBT genes, enhancers also play important roles in regulating gene transcription in more specific ways. Whether a gene is activated during pre-MBT or MBT depends on the combined regulatory input of enhancers and the core promoter.

A conserved mechanism of ZGA during MBT?

Although this general model is proposed for *Drosophila*, it might also be conserved for other vertebrate systems with MBT. As mentioned above, functional studies for TBP in zebrafish and *Xenopus* have shown that manipulating the concentration of TBP only affects a small portion of early genes, rather than global ZGA (Prioleau et al., 1994; Müller et al., 2001). In addition, knockdown of TBP in zebrafish also compromises the degradation of maternal RNA (Ferg et al., 2007). This is similar to my model in *Drosophila*, in which the zygotic miRNAs transcribed in pre-MBT embryos, which are more sensitive to TBP, mediate global maternal RNA degradation. Indeed, in zebrafish, there is a cluster of zygotic miRNAs that plays a similar role in mediating maternal RNA degradation (Giraldez et al., 2006). Although there is H3K4me3 in the early zebrafish embryos, genes involved in DV patterning and cell fate determination have less H3K4me3 than other zygotic genes (Lindeman et al., 2011). This is also similar to my observation in *Drosophila* that the early body patterning genes have fewer histone modifications. Taken together, these studies imply that the roles of TBP and core promoters may be conserved in regulating the early transcription in vertebrate systems, at least for zebrafish and *Xenopus*.

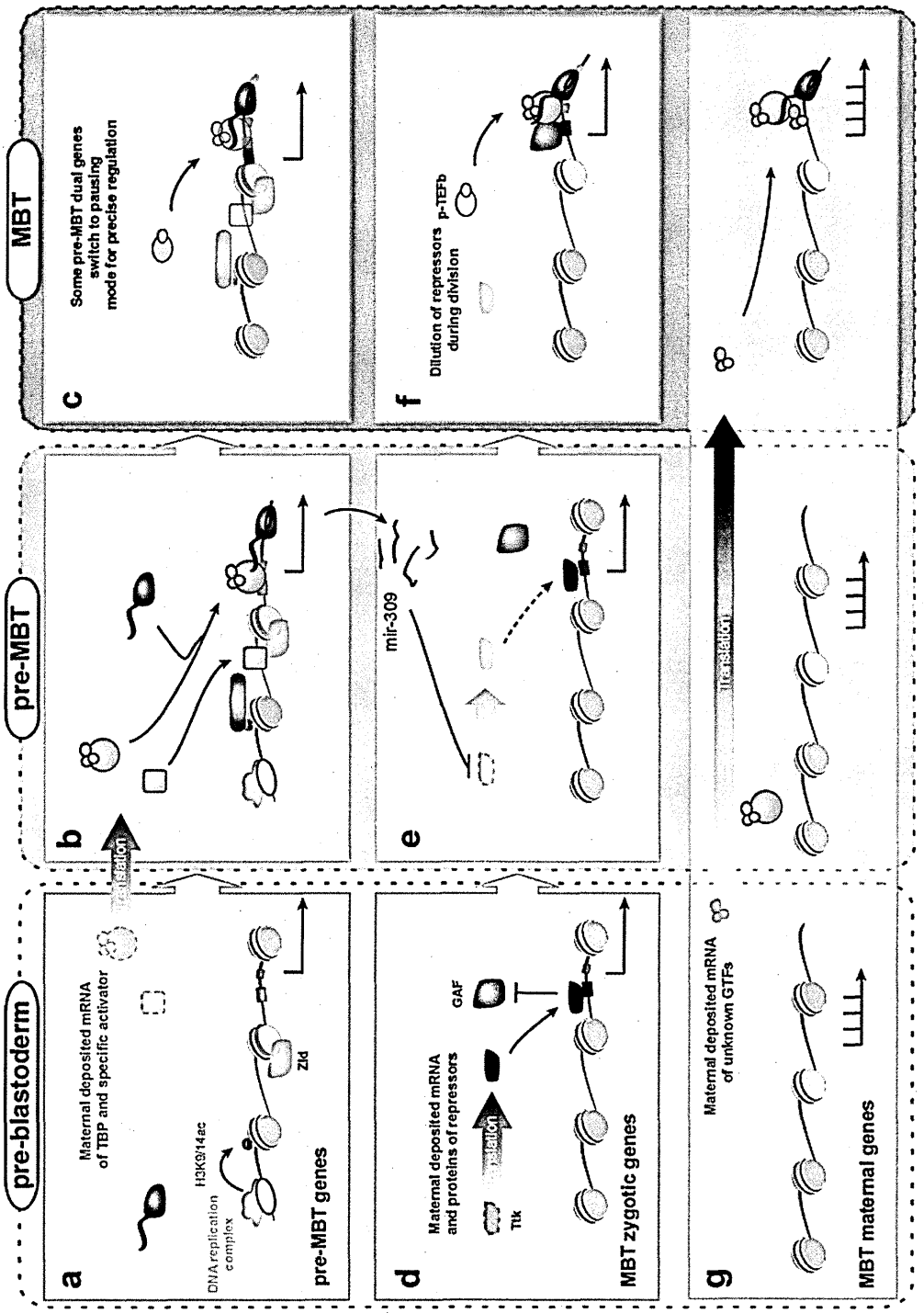


Figure 4-1. Model for timing the differential onsets of zygotic genes during *Drosophila* MZT.
Initially, chromatin is loosely packed and accessible to transcription factors. **a,** The lack of transcription is due to the absence of TBP proteins, which are later translated by maternal mRNA in pre-MBT embryos (**b**). **c,** Some pre-MBT genes that need precise regulation can switch to pausing mode after the initial quick divisions have finished. **d,** Due to the repression of maternal deposited repressors, MBT zygotic genes fail to recruit GAF and Pol II in the pre-MBT embryos. **e,** pre-MBT transcribed mir-309 mediates mRNA degradation, which stops the supply of newly translated repressor. **f,** The dilution of repression during division makes the promoters of MBT zygotic genes available for GAF and Pol II. **g,** The activation of MBT maternal genes may require some unknown GTFs, which start to function in the MBT embryos.

4.2 Conclusion and Future perspectives

In this study, I have successfully established a system to perform ChIP-Seq and nucleosome mappings for early *Drosophila* embryos with a limited amount of material. Based on this system, I systematically performed ChIP-Seq and obtained high quality genome-wide profiles for general transcription factors and histone modifications, as well as nucleosome occupancy, on *Drosophila* pre-MBT, MBT, and post-MBT embryos. In agreement with our histologic observations, the ChIP-Seq and MNase-Seq results suggest that 1) the initial chromatin is loosely packed by nucleosomes and is highly accessible to various DNA binding factors, 2) widespread paused Pol II is established during MBT without pre-patterned histone modifications, and 3) the PcG-mediated silencing is re-established at the end of cellularization. Furthermore, by comparing the core promoter motifs of the pre-MBT genes with those of later activated genes, we found that the pre-MBT genes are highly enriched for the TATA-box, and I propose a possible model to explain how these different core promoters may time the gene activation during early embryogenesis.

The ChIP-Seq system I built for the early embryo is no doubt valuable for exploring many questions about the early chromatin state. For example, besides the histone modifications I test in this study, can other histone modification markers serve as epigenetic markers for instructing early transcription? How do those early developmental regulators, like Bcd and Zld regulate the early chromatin accessibility? And how do repressors and GAF regulate the differential activations of pre-MBT and MBT genes?

Our findings regarding the formation of chromatin signatures during MBT and the differential usage of core promoters also generate many new questions. Here are some general directions that I think are worthwhile to design follow-up experiments to explore in greater depth.

4.2.1 Improving the understanding of chromatin organization during embryogenesis with genomic approaches

So far, most *in vivo* genomic studies in *Drosophila* use whole embryos (The modENCODE Consortium et al., 2010) despite the tissue diversity, to understand the chromatin organization in a developmental context. These accumulating data are important for annotating the genome, and they provide informative insights about the dynamic changes in chromatin organization during development. However, sometimes it is hard to

interpret and connect those dynamic changes to certain parts of organogenesis or other development events. In this study, for example, the results show differential deposition of H3K27me3 along the AP axis during post-MBT, while the whole embryo ChIP-Seq can only tell us when the H3K27me3 pattern is established. We cannot distinguish whether additional peaks found in later embryos are added to H3K27me3 high-staining cells or newly formed in H3K27me3 weak-staining cells. By combining fluorescence-activated cell sorting (FACS) with cell-specific GFP marker (Gaertner et al., 2012) or tag-protein followed fluorescence-staining (Bonn et al., 2012), it is now feasible to understand the dynamic changes in chromatin structure at cell type–specific resolution during embryogenesis. Nevertheless, for certain types of cells, like the pole cells, whose population is very limited, FACS would be very costly and time consuming. In that case, the strategy of using biotin labeling at nuclear membrane proteins, recently developed by the Henikoff lab (Deal and Henikoff, 2010), would be helpful.

Moreover, instead of being linearly packed into nuclei, the chromatin is highly organized into many sub-nuclear domains, in which different loci with similar transcriptional activity contact each other to form the long-range looping structures. Rather than detecting the interactions between a few loci by chromatin FISH, a recent genome-wide chromosome conformation capture study on *Drosophila* embryonic nuclei (Sexton et al., 2012) depicts a beautiful chromatin interaction map for numerous physical contact loci and lays the foundation for detailed studies of chromatin structure and function. If we can combine this technique with the cell-specific ChIP and apply this method to early embryos, it will help us answer questions like when do those sub-nuclear structures form during development? How are those structures formed in response to signals? And what is the connection between those structures and cell fate determination during development?

With the dropping costs of sequencing, I imagine that more and more laboratories will join this field to address these questions. Hopefully, massive amounts of data generated through the high throughput technology and strategies mentioned above will help us to uncover a comprehensive understanding about chromatin organization during embryogenesis.

4.2.2 Expanding the knowledge of the roles of core promoters and pre-MBT genes in early development

Although our results imply that the promoter might play roles in timing

transcription during early embryogenesis, this hypothesis still needs to be confirmed by more functional studies, such as inserting the core promoter of the pre-MBT genes to the later activated genes to see whether it can cause the premature transcription of the later activated genes in the early embryo. In addition, to figure out which GAFs recognize different core promoters and are involved in the timing of early transcription, we can perform RNAi screens for GTFs to test whether any proteins only compromise the activation of pre-MBT genes or later activated genes *in vivo*. Moreover, biochemical tools can also be used to identify factors involved in developmental gene regulation. One strategy is using the immobilized DNA template assay (Johnson et al., 2004) to identify which GTFs are only pulled down with certain core promoter sequences.

In addition to the transcriptional regulation, the roles of those pre-MBT genes I identified are also valuable for studying. Although 35 of these genes have been identified and studied through genetic screening, more than half of the pre-MBT genes have not been studied previously. For example, among the pre-MBT genes with unknown functions, there are 7 genes that contain a common unknown function domain, DUF733, which is also found in two other pre-MBT genes, *SNCF* and *Halo* (Bonneaud et al., 2003). This unknown function domain is well conserved in all 12 *Drosophila* species and only found in the pre-MBT genes. What is the role of those unknown proteins and this conserved domain? What are the transcriptional patterns of those genes, and where are they located in the cells? Since most known pre-MBT genes are transcription factors, are they also involved in transcriptional regulation?

Finally, it would be interesting to test whether the role of the core promoters in regulating early transcription is conserved in other vertebrate systems. For zebrafish, although there is some TBP knockdown evidence, the specific regulatory action of TBP and focused promoters still unknown. How do these core promoters correlate with H3K4me3, and together to regulate early gene expression? For mammals, which do not have MBT, how do those core promoters regulate the early gene activation during MZT?

I believe that over next several years, these studies will assist us in elucidating the roles of the core promoters and GTFs in setting up the gene regulatory networks in *Drosophila* embryogenesis. These studies might reveal more unexpected functions of core promoters and GTFs in other developmental events or pathogenesis processes, not only *Drosophila*, but also in humans.

Appendix

Data analysis was performed with the help of Julia Zeitlinger, Jeff Johnston and Sam Meier. All genomic data is available upon request. This Appendix summarizes all bioinformatics analysis performed by Jeff Johnston and Sam Meier.

Alignment of ChIP-Seq and MNase-Seq data

All sequencing reads were aligned to the UCSC *Drosophila melanogaster* dm3 genome with Bowtie v0.12.8 (Langmead et al., 2009) using the following parameters:

-k 1 –m 1 –l 40 –n 2 –best –strata

The MBT MNase-Seq library was paired-end sequenced and alignment was performed with an allowable insert size of 47 to 297 bp. After alignment, single-end reads were extended to the estimated insert size of the library as determined by a Bioanalyzer. To identify alignment and amplification artifacts, custom R scripts were used to analyze the aligned reads of all single-end libraries with more than 10 duplicates (defined as having the same chromosome, start and strand values). These "stacks" of identical reads were removed unless a corresponding number of reads were present on the opposite strand approximately one fragment length away in the 3' direction. For all libraries, genome-wide coverage was calculated by assigning an integer score to each genomic coordinate representing the number of extended reads that overlapped that location.

Analysis of RNA-Seq expression data

To obtain gene expression measurements at different mitotic cycles, we downloaded single-embryo RNA-Seq datasets from http://eisenlab.org/dosage/ (Lott et al., 2011). One female and one male replicate were downloaded for mitotic cycles 10, 11, 12, 13 and 14 (A-D). The male and female datasets were combined for each mitotic cycle. In addition, we downloaded 4-6 h and 6-8 h staged whole-embryo RNA-Seq datasets (Graveley et al., 2011). We processed all samples using TopHat v2.0.4 (Trapnell et al., 2009) by aligning against the FlyBase r5.47 genome and its corresponding gene annotations using the following parameters:

Single-embryo samples (40bp reads):

-G fb-r5.47.gtf –I 5000 –segment-length 20 fb547_genome

4-6 h and 6-8 h embryo samples (75bp reads):

-G fb-r5.47.gtf –I 5000 –segment-length 37 fb547_genome

Next, we used the cuffdiff tool from Cufflinks v2.0.2 (Trapnell et al., 2010) to obtain gene expression values (RPKMs) for all samples using the following non-default parameters:

-u –b fb547_genome.fa fb-r5.47.gtf

Calculating Pol II enrichments and stalling indexes

For the four Pol II replicates in the pre-MBT embryo and the three Pol II replicates in the MBT embryo, enrichment ratios were calculated for the TSS region, a region immediately downstream of the TSS, and the transcription unit (TU) region of each unique FlyBase r5.47 transcript. The TSS region was defined as the first 200bp of the transcript, and the downstream TSS region was defined as the 200bp following the TSS region (201 to 400bp). The TU region was defined as 400bp downstream of the TSS to the end of the transcript. For transcripts less than 600bp in length, the TU region was defined as the entire transcript. Total signal for each region was found for each Pol II and WCE sample. Enrichment in each region was calculated after normalizing for both fragment length and total read count:

Enrichment = (IP Signal / (IP read count × IP fragment length)) /
(WCE Signal / (WCE read count × WCE fragment length))

The stalling index for each gene was defined as: $\log_2$ Pol $II_{(TSS)} - \log_2$ Pol $II_{(Downstream\ TSS)}$ after flooring both Pol II enrichment values at 1 (background). Stalling indexes for all replicates were averaged.

Identification and classification of pre-MBT genes

To identify genes bound by Pol II in the pre-MBT embryo, we first identified all transcripts with Pol II_{TSS} enrichment two-fold above WCE in all four replicates. To ensure that these enrichments were due to high Pol II signal, we also required the Pol II signal

portion of the enrichment calculation (the numerator in the above equation) to be in the 99th percentile of all transcripts in all four replicates.

Manual inspection of some of these transcripts showed that the Pol II signal originated from a different gene's TSS. To eliminate these false positives, we used MACS to identify peaks in our best pre-MBT TBP sample and manually examined all Pol II-enriched pre-MBT transcripts that did not have a detected TBP peak within 500bp of the TSS. We used the default parameters of MACS v2.0.10.20120703 (Zhang et al., 2008), specifying only the preset alignable genome size for *Drosophila melanogaster* using the "-g dm" argument. This identified 12 transcripts in which the Pol II signal did not appear to originate from the TSS. These transcripts were removed from our pre-MBT list and are marked as "rejected pre-MBT genes" in the Supplemental Spreadsheet.

We next checked for possible pre-MBT genes with missing or mis-annotated transcription start sites. To do this, we used MACS to call peaks on all four of our pre-MBT Pol II samples using the same default parameters as described above. We then identified all regions that were called as peaks in at least two of the four replicates. These regions were assigned to the nearest gene within 5kb and all regions assigned to a gene not already considered a pre-MBT gene were manually examined. This revealed ten possible additional pre-MBT genes where the Pol II signal originated from an un-annotated transcription start site. As all of these genes also had at least some TBP signal upstream of the Pol II signal, we defined custom transcript entries for these genes by setting the transcript start site to 19bp downstream of the location of the maximum TBP signal. To ensure these custom transcripts met our existing enrichment criteria, we performed the same calculations as described above in the **Calculating Pol II enrichments** section. All ten of the custom transcripts were sufficiently enriched in Pol II and were added to our pre-MBT gene list.

We classified the 116 pre-MBT genes into three groups. First, the "paused" group was defined as those pre-MBT genes having a mean (among all four replicates) Pol II_{TU} ratio less than 1. The "dual" group was defined as any pre-MBT gene not in the paused group that had Pol II_{TSS} enrichment in the top 20% of all genes in 6-8hr Mef2-positive muscle cells. The remaining pre-MBT genes were classified as the "not paused" group.

Identification and classification of MBT genes

To identify genes bound by Pol II in the MBT embryo, we selected all transcripts with Pol II_{TSS} enrichment at least two-fold above WCE in all three replicates. If multiple transcripts for the same gene met these criteria, we selected the one with the highest Pol II_{TSS} signal (breaking ties using the mean Pol II_{TU} enrichment).

MBT genes were classified into three groups using gene expression values calculated from previously published single-embryo RNA-Seq experiments (see **Analysis of RNA-Seq expression data** section). We classified as "maternal" all MBT genes with an RPKM of at least 1 in mitotic cycle 10. We classified as "MBT active" all non-maternal MBT genes with an RPKM of at least 5 in mitotic cycle 14D. The remaining MBT genes were classified as "MBT paused."

Normalization of reads and enrichment values

For the heatmap in Figure 3-2, enrichment values were first calculated in a 100bp sliding window across all samples. Replicates were combined by taking the minimum enrichment value at each base. Samples were then independently normalized by defining "minimum" as an enrichment value of 1 (background) and "maximum" as the 99^{th} percentile enrichment value encountered among all displayed bases. As there was no significant ChIP enrichment in the pre-MBT H3K4me3 sample, it was normalized to the maximum enrichment value of the MBT H3K4me3 sample to avoid amplifying noise.

For Figure 3-5 and 3-6, both read counts and enrichment values for each sample were independently scaled by dividing the values at each base by the maximum value encountered among the displayed genes or gene groups across stages after normalizing for both read count and fragment size differences.

Metapeak analysis for H3K27me3 and Pc ChIP-Seq

Using supplemental table 17 from Schuettengruber et al.(Schuettengruber et al., 2009), the 441 regions were selected based on Ph ChIP-chip enrichment (p-value < 0.0001). The regions were then aligned at their midpoints and extended by 80kb in both directions. Average region graphs were constructed showing the average enrichment value at each base for three H3K27me3 samples. The enrichment for each sample was defined by

dividing read-count normalized IP signal by the read-count normalized whole cell extract control signal.

Promoter element annotations

Sequences surrounding all FlyBase r5.47 transcription start sites (plus our ten additional custom pre-MBT transcripts) were scanned for the core promoter elements listed in [Promoter element supplemental table]. A core promoter element was scored as present if no mismatch was found within a specified window relative to the transcription start site. For Zelda, we also counted the number of motifs found in each transcript's window.

Promoter element enrichments

For each group of transcripts analyzed for promoter element composition, an enrichment and a p-value were calculated for each promoter element. Enrichment was calculated as follows, where G is the group of transcripts tested and PE is a particular promoter element:

Observed = (# of transcripts in G with element PE) / (# of transcripts in G)
Expected = (# of transcripts in the genome with element PE) /
(# of transcripts in the genome)
Enrichment = Observed / Expected

For enrichment values less than one, the negative reciprocal of the enrichment value was used (indicating depletion instead of enrichment). To calculate a p-value for the observed frequency of each promoter element in each group of transcripts, a Fisher test was performed. Enrichments and depletions with a p-value less than 0.05 (after correcting for multiple testing with the Benjamini & Hochberg method) were deemed significant.

For Zelda, we calculated enrichment and p-values via random sampling. Enrichment values for each group of transcripts were calculated by dividing the number of Zelda sites per transcript in each group by the average number of Zelda sites per transcript in the genome. To calculate p-values for each group of transcripts, we randomly selected an equal number of transcripts from the entire genome 10,000 times and calculated the enrichment value for each random sample. The p-value was then calculated as the portion of random samples with higher Zelda enrichment than the transcript group.

Promoter type enrichments

To calculate enrichment of gene groups for promoter types, we first downloaded the promoter classifications from two sources (Ni et al., 2010; Hoskins et al., 2011). The Ohler Lab classified promoters into Broad, Broad with Peak, and Peaked; the Celniker Lab classified promoters into Broad, Unclassified and Peaked. Using FlyBase r5.47 transcripts, we assigned each transcript a promoter type from each dataset if a classified promoter was within 10bp of the transcript's start site. This yielded 12,039 annotated transcripts for the Celniker Lab data and 4,457 annotated transcripts for the Ohler Lab data. For each gene group analyzed, we calculated enrichments and p-values using the same method as the promoter element enrichment analysis described above.

Table A-1. List of core promoter elements, their consensus sequences and relative position to the TSS

Motif Name	IUPAC Consensus	Directional	Window	Transcript count	Reference
DPE	KCGGTTSK	Yes	0 to 75	537	(Burke and Kadonaga, 1996)
DRE	WATCGATW	Yes	-100 to 0	2,111	(Hochheimer et al., 2002)
GAGA	GAGA	No	-100 to 0	9,559	(Stark et al., 2007)
Inr	TCAKTY	Yes	-50 to 50	5,965	(Smale and Baltimore, 1989)
MTE	CSARCSSA	Yes	0 to 30	212	(Lim, 2004)
PB	KCGRWCG	Yes	-50 to 100	2,093	(Hendrix et al., 2008)
TATA	STATAWAWR	Yes	-100 to 0	1,503	(Goldberg, 1979)
Motif1	YGGTCACACTR	Yes	-100 to 50	609	(Ohler et al., 2002)
Motif6	YRGTATWTTY	Yes	-150 to 25	840	(Ohler et al., 2002)
Motif7	CAKCNCTR	Yes	-100 to 50	2,190	(Ohler et al., 2002)
Zelda	YAGGTAR	No	-2000 to 0	9,798	(Bosch et al., 2006; Liang et al., 2008)

de novo motif discovery

Fasta sequence files were generated for each of the classified Pre-MBT and MBT groups, based on regions +/- 100bp surrounding the FlyBase r5.47 transcription start sites (plus our ten additional custom pre-MBT transcripts). Using MEME v4.8.1 (Bailey et al., 2009) the fasta files were processed using the following parameters:

-mod zoops -dna -nmotifs 50 -revcomp -maxw 12 -maxsize 5000000 -oc meme/

The resulting motifs were then compared against the TRANSFAC 2011.4 database and the promoter element table listed above using the TOMTOM tool also from the MEME suite.

ImaGO earliest annotated expression analysis

To calculate enrichment of gene groups for expression in specific stages, we downloaded the Berkeley *Drosophila* Genome Project *in situ* expression database from http://insitu.fruitfly.org/ (Tomancak et al., 2002; 2007). We removed the "maternal" and "no staining" annotation entries and then removed all but the first (in stage order) annotation of each gene. For each gene group analyzed, we calculated enrichments and p-values using the same method as the promoter element enrichment analysis described above.

Analysis of gene expression over time

For all gene groups plotted in Figure 2B and S4, we first removed any genes with evidence of maternally deposited mRNA using the following criteria:

- RPKM expression > 16 in mitotic cycle 10, or
- Maternal expression at least two-fold above zygotic expression in mitotic cycle 10 (F10 sample in Dataset S1 of (Lott et al., 2011))

References

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. *13*, 720–731.

Aggarwal, B.D., and Calvi, B.R. (2004). Chromatin regulates origin activity in Drosophila follicle cells. Nature *430*, 372–376.

Akkers, R.C., van Heeringen, S.J., Jacobi, U.G., Janssen-Megens, E.M., Françoijs, K.-J., Stunnenberg, H.G., and Veenstra, G.J.C. (2009). A Hierarchy of H3K4me3 and H3K27me3 Acquisition in Spatial Gene Regulation in Xenopus Embryos. Developmental Cell *17*, 425–434.

Alvarado, A. (2003). The freshwater planarian Schmidtea mediterranea: embryogenesis, stem cells and regeneration. Current Opinion in Genetics & Development *13*, 438–444.

Amir-Zilberstein, L., Amir-Zilberstein, L., Ainbinder, E., Ainbinder, E., Toube, L., Toube, L., Yamaguchi, Y., Yamaguchi, Y., Handa, H., Handa, H., et al. (2007). Differential regulation of NF-kappaB by elongation factors is determined by core promoter type. Mol. Cell. Biol. *27*, 5246–5259.

Amleh, A., Nair, S.J., Sun, J., Sutherland, A., Hasty, P., and Li, R. (2009). Mouse cofactor of BRCA1 (Cobra1) is required for early embryogenesis. PLoS ONE *4*, e5034.

Anderson, K.V., and Lengyel, J.A. (1979). Rates of synthesis of major classes of RNA in Drosophila embryos. Dev. Biol. *70*, 217–231.

Anderson, K.V., and Lengyel, J.A. (1980). Changing rates of histone mRNA synthesis and turnover in Drosophila embryos. Cell *21*, 717–727.

Arking, R., and Parente, A. (1980). Effects of RNA inhibitors on the development of Drosophila embryos permeabilized by a new technique. J. Exp. Zool. *212*, 183–194.

Aso, T., Juven-Gershon, T., Conaway, J.W., Kadonaga, J.T., and Conaway, R.C. (1994). Role of core promoter structure in assembly of the RNA polymerase II preinitiation complex. A common pathway for formation of preinitiation intermediates at many TATA and TATA-less promoters. J. Biol. Chem. *269*, 26575–26583.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research *37*, W202–W208.

Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., and Cavalli, G. (2011). Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. Cell *144*, 214–226.

Bao, S., Miyoshi, N., Okamoto, I., Jenuwein, T., Heard, E., and Azim Surani, M. (2005). Initiation of epigenetic reprogramming of the X chromosome in somatic nuclei transplanted to a mouse oocyte. EMBO Rep *6*, 748–754.

Bashirullah, A., Halsell, S.R., Cooperstock, R.L., Kloc, M., Karaiskakis, A., Fisher, W.W., Fu, W., Hamilton, J.K., Etkin, L.D., and Lipshitz, H.D. (1999). Joint action of two RNA degradation pathways controls the timing of maternal transcript elimination at the midblastula transition in Drosophila melanogaster. The EMBO Journal *18*, 2610–2620.

Beisel, C., and Paro, R. (2011). Silencing chromatin: comparing modes and mechanisms. Nat. Rev. Genet. *12*, 123–135.

Bell, O., Tiwari, V.K., Thomä, N.H., and Schübeler, D. (2011). Determinants and dynamics of genome accessibility. Nat. Rev. Genet. *12*, 554–564.

Benoit, B., He, C.H., Zhang, F., Votruba, S.M., Tadros, W., Westwood, J.T., Smibert, C.A., Lipshitz, H.D., and Theurkauf, W.E. (2009). An essential role for the RNA-binding protein Smaug during the Drosophila maternal-to-zygotic transition. Development *136*, 923–932.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell *125*, 315–326.

Boettiger, A.N., and Levine, M. (2009). Synchronous and stochastic patterns of gene activation in the Drosophila embryo. Science *325*, 471–473.

Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczyński, B., Riddell, A., and Furlong, E.E.M. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat Genet *44*, 148–156.

Bonneaud, N., Savare, J., Berta, P., and Girard, F. (2003). SNCF, a SoxNeuro interacting protein, defines a novel protein family in Drosophila melanogaster. Gene *319*, 33–41.

Bosch, ten, J.R., Benavides, J.A., and Cline, T.W. (2006). The TAGteam DNA motif controls the timing of Drosophila pre-blastoderm transcription. Development *133*, 1967–1977.

Bothma, J.P., Magliocco, J., and Levine, M. (2011). The Snail Repressor Inhibits Release, Not Elongation, of Paused Pol II in the Drosophila Embryo. Current Biology.

Brown, J.L., Sonoda, S., Ueda, H., Scott, M.P., and Wu, C. (1991). Repression of the Drosophila fushi tarazu (ftz) segmentation gene. The EMBO Journal *10*, 665–674.

Brykczynska, U., Hisano, M., Erkek, S., Ramos, L., Oakeley, E.J., Roloff, T.C., Beisel, C., Schübeler, D., Stadler, M.B., and Peters, A.H.F.M. (2010). Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. Nature Structural & Molecular Biology *17*, 679–687.

Buchenau, P., Hodgson, J., Strutt, H., and Arndt-Jovin, D.J. (1998). The distribution of polycomb-group proteins during cell division and development in Drosophila embryos: impact on models for silencing. J. Cell Biol. *141*, 469–481.

Bultman, S.J., Gebuhr, T.C., Pan, H., Svoboda, P., Schultz, R.M., and Magnuson, T. (2006). Maternal BRG1 regulates zygotic genome activation in the mouse. Genes & Development *20*, 1744–1754.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Mol. Cell *36*, 541–546.

Burgess, R.J., Zhou, H., Han, J., and Zhang, Z. (2010). A role for Gcn5 in replication-coupled nucleosome assembly. Mol. Cell *37*, 469–480.

Burke, T.W., and Kadonaga, J.T. (1996). Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *10*, 711–724.

Bushati, N., Stark, A., Brennecke, J., and Cohen, S.M. (2008). Temporal Reciprocity of miRNAs and Their Targets during the Maternal-to-Zygotic Transition in Drosophila. Current Biology *18*, 501–506.

Chambeyron, S. (2005). Nuclear re-organisation of the Hoxb complex during mouse embryonic development. Development *132*, 2215–2223.

Cheng, B., Li, T., Rahl, P.B., Adamson, T.E., Loudas, N.B., Guo, J., Varzavand, K., Cooper, J.J., Hu, X., Gnatt, A., et al. (2012). Functional association of Gdown1 with RNA polymerase II poised on human genes. Mol. Cell *45*, 38–50.

Cheutin, T., and Cavalli, G. (2012). Progressive polycomb assembly on H3K27me3 compartments generates polycomb bodies with developmentally regulated motion. PLoS Genet *8*, e1002465.

Chopra, V.S., Hendrix, D.A., Core, L.J., Tsui, C., Lis, J.T., and Levine, M. (2011). The Polycomb Group Mutant esc Leads to Augmented Levels of Paused Pol II in the Drosophila Embryo. Mol. Cell *42*, 837–844.

Chopra, V.S., Hong, J.-W., and Levine, M. (2009). Regulation of Hox gene activity by transcriptional elongation in Drosophila. Curr. Biol. *19*, 688–693.

Corden, J.L. (1990). Tails of RNA polymerase II. Trends in Biochemical Sciences *15*, 383–387.

Cui, K., Zang, C., Roh, T.Y., Schones, D.E., and Childs, R.W. (2009). Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell Stem Cell.

De Renzis, S., Elemento, O., Tavazoie, S., and Wieschaus, E.F. (2007). Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo. PLoS Biol *5*, e117.

Deal, R.B., and Henikoff, S. (2010). A Simple Method for Gene Expression and Chromatin Profiling of Individual Cell Types within a Tissue. Developmental Cell *18*, 1030–1040.

Denell, R.E. (1978). Homoeosis in Drosophila. II. a Genetic Analysis of Polycomb. Genetics *90*, 277–289.

Dernburg, A.F., Broman, K.W., Fung, J.C., Marshall, W.F., Philips, J., Agard, D.A., and Sedat, J.W. (1996). Perturbation of nuclear architecture by long-distance chromosome interactions. Cell *85*, 745–759.

Déjardin, J., and Cavalli, G. (2004). Chromatin inheritance upon Zeste-mediated Brahma

recruitment at a minimal cellular memory module. The EMBO Journal *23*, 857–868.

Doane, W.W. (1960). Completion of meiosis in uninseminated eggs of Drosophila melanogaster. Science *132*, 677–678.

Driever, W., Driever, W., Nüsslein-Volhard, C., and Nüsslein-Volhard, C. (1988). A gradient of bicoid protein in Drosophila embryos. Cell *54*, 83–93.

Eddy, J., Vallur, A.C., Varma, S., Liu, H., Reinhold, W.C., Pommier, Y., and Maizels, N. (2011). G4 motifs correlate with promoter-proximal transcriptional pausing in human genes. Nucleic Acids Research *39*, 4975–4983.

Edgar, B.A., and Schubiger, G. (1986). Parameters controlling transcriptional activation during early Drosophila development. Cell *44*, 871–877.

ensembl, T. http://useast.ensembl.org/Homo_sapiens/Gene/Compara_Tree?db=core;g=ENSG00000165632;r=10:7860467-8058590.

Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. Nature *421*, 448–453.

Ferg, M., Sanges, R., Gehrig, J., Kiss, J., Bauer, M., Lovas, A., Szabo, M., Yang, L., Straehle, U., Pankratz, M.J., et al. (2007). The TATA-binding protein regulates maternal mRNA degradation and differential zygotic transcription in zebrafish. The EMBO Journal *26*, 3945–3956.

FitzGerald, P.C., Sturgill, D., Shyakhtenko, A., Oliver, B., and Vinson, C. (2006). Comparative genomics of Drosophila and human core promoters. Genome Biol. *7*, R53.

Foe, V.E., Odell, G.M., and Edgar, B.A. (1993). Mitosis and Morphogenesis in the Drosophila Embryo. In The Development of Drosophila Melanogaster, M. Bate, and A.M. Arias, eds. (Cold Spring Harbor Laboratory Pr), pp. 149–300.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. Science *306*, 1574–1577.

Fung, J.C., Marshall, W.F., Dernburg, A., Agard, D.A., and Sedat, J.W. (1998). Homologous chromosome pairing in Drosophila melanogaster proceeds through multiple independent initiations. J. Cell Biol. *141*, 5–20.

Gaertner, B., Johnston, J., Kai, C., Wallaschek, N., Paulson, A., Garruss, A.S., Gaudenz, K., De Kumar, B., Krumlauf, R., and Zeitlinger, J. (2012). Poised RNA Polymerase II Changes over Developmental Time and Prepares Genes for Future Expression. Cell Rep.

Garcia-Fernàndez, J. (2005). The genesis and evolution of homeobox gene clusters. Nat. Rev. Genet. *6*, 881–892.

Gemkow, M.J., Verveer, P.J., and Arndt-Jovin, D.J. (1998). Homologous association of the Bithorax-Complex during embryogenesis: consequences for transvection in Drosophila melanogaster. Development *125*, 4541–4552.

Gilchrist, D.A., Santos, G.D., Fargo, D.C., Xie, B., Gao, Y., Li, L., and Adelman, K.

(2010). Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. Cell *143*, 540–551.

Gilmour, D.S., and Lis, J.T. (1985). In vivo interactions of RNA polymerase II with genes of Drosophila melanogaster. Mol. Cell. Biol. *5*, 2009–2018.

Gilmour, D.S., and Lis, J.T. (1986). RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. Mol. Cell. Biol. *6*, 3984–3989.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75–79.

Giresi, P.G., Kim, J., McDaniell, R.M., Iyer, V.R., and Lieb, J.D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Research *17*, 877–885.

Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell.

Goldberg, M.L. (1979). Sequence analysis of Drosophila histone genes.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., et al. (2011). The developmental transcriptome of Drosophila melanogaster. Nature *471*, 473–479.

Grewal, S.I.S., and Jia, S. (2007). Heterochromatin revisited. Nat. Rev. Genet. *8*, 35–46.

Gu, W., Powell, W., Mote, J., and Reines, D. (1993). Nascent RNA cleavage by arrested RNA polymerase II does not require upstream translocation of the elongation complex on DNA. J. Biol. Chem. *268*, 25604–25616.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77–88.

Gurdon, J.B. (1962). The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. J Embryol Exp Morphol *10*, 622–640.

Gurdon, J.B., and Melton, D.A. (2008). Nuclear reprogramming in cells. Science *322*, 1811–1815.

Gutzeit, H.O. (1980). Expression of the zygotic genome in blastoderm stage embryos ofDrosophila: Analysis of a specific protein. Archiv Für Entwickelungsmechanik Der Organismen *188*, 153–156.

Hajduskova, M., Ahier, A., Daniele, T., and Jarriault, S. (2011). Cell plasticity in Caenorhabditis elegans: From induced to natural cell reprogramming. Genesis *50*, 1–17.

Hammoud, S.S., Nix, D.A., Zhang, H., Purwar, J., Carrell, D.T., and Cairns, B.R. (2009). Distinctive chromatin in human sperm packages genes for embryo development. Nature *460*, 473–478.

Haynie, J.L. (1983). The maternal and zygotic roles of the gene Polycomb in embryonic determination in Drosophila melanogaster. Dev. Biol. *100*, 399–411.

Heifetz, Y., Yu, J., and Wolfner, M.F. (2001). Ovulation triggers activation of Drosophila oocytes. Dev. Biol. *234*, 416–424.

Hendrix, D.A., Hong, J.-W., Zeitlinger, J., Rokhsar, D.S., and Levine, M.S. (2008). Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. Proceedings of the National Academy of Sciences *105*, 7762–7767.

Hiraoka, Y., Dernburg, A.F., Parmelee, S.J., Rykowski, M.C., Agard, D.A., and Sedat, J.W. (1993). The onset of homologous chromosome pairing during Drosophila melanogaster embryogenesis. J. Cell Biol. *120*, 591–600.

Hochheimer, A., Zhou, S., Zheng, S., Holmes, M.C., and Tjian, R. (2002). TRF2 associates with DREF and directs promoter-selective gene expression in Drosophila. Nature *420*, 439–445.

Holliday, R. (1994). Epigenetics: an overview. Dev. Genet. *15*, 453–457.

Hong, J.-W., Hendrix, D.A., Papatsenko, D., and Levine, M.S. (2008). How the Dorsal gradient works: insights from postgenome technologies. Proc Natl Acad Sci USA *105*, 20072–20076.

Horner, V.L., and Wolfner, M.F. (2008). Mechanical stimulation by osmotic and hydrostatic pressure activates Drosophila oocytes in vitro in a calcium-dependent manner. Dev. Biol. *316*, 100–109.

Hornung, G., Lindeman, L.C., Bar-Ziv, R., Andersen, I.S., Rosin, D., Reiner, A.H., Tokuriki, N., Li, N., Tawfik, D.S., Aanes, H., et al. (2012). Noise-mean relationship in mutated promoters. Genome Research.

Hoskins, R.A., Hoskins, R.A., Landolin, J.M., Landolin, J.M., Brown, J.B., Brown, J.B., Sandler, J.E., Sandler, J.E., Takahashi, H., Takahashi, H., et al. (2011). Genome-wide analysis of promoter architecture in Drosophila melanogaster. Genome Research *21*, 182–192.

Ingham, P.W. (1983). Differential expression of bithorax complex genes in the absence of the extra sex combs and trithorax genes. Nature *306*, 591–593.

Johnson, K.M., Wang, J., Smallwood, A., and Carey, M. (2004). The immobilized template assay for measuring cooperativity in eukaryotic transcription complex assembly. Meth. Enzymol. *380*, 207–219.

Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev. Biol. *339*, 225–229.

Jürgens, G. (1985). A group of genes controlling the spatial expression of the bithorax complex in Drosophila. Nature *316*, 153–155.

Jürgens, G., Wieschaus, E., N sslein-Volhard, C., and Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle in Drosophila melanogaster. Development Genes

and Evolution *193*, 283–295.

Kadauke, S., and Blobel, G.A. (2009). Chromatin loops in gene regulation. BBA - Gene Regulatory Mechanisms *1789*, 17–25.

Kane, D.A., and Kimmel, C.B. (1993). The zebrafish midblastula transition. Development *119*, 447–456.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., et al. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362–366.

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T., et al. (2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature *471*, 480–485.

Kim, N., and Jinks-Robertson, S. (2012). Transcription as a source of genome instability. Nat. Rev. Genet. *13*, 204–214.

Kimelman, D., Kirschner, M., and Scherson, T. (1987). The events of the midblastula transition in Xenopus are regulated by changes in the cell cycle. Cell *48*, 399–407.

Konrad, K.D., Goralski, T.J., and Mahowald, A.P. (1988). Developmental genetics of the gastrulation defective locus in Drosophila melanogaster. Dev. Biol. *127*, 133–142.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868–871.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705.

Lamb, M.M., and Laird, C.D. (1976). Increase in nuclear poly(A)-containing RNA at syncytial blastoderm in Drosophila melanogaster embryos. Dev. Biol. *52*, 31–42.

Lan, F., Bayliss, P.E., Rinn, J.L., Whetstine, J.R., Wang, J.K., Chen, S., Iwase, S., Alpatov, R., Issaeva, I., Canaani, E., et al. (2007). A histone H3 lysine 27 demethylase regulates animal posterior development. Nature *449*, 689–694.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lee, C., Li, X., Hechmer, A., Eisen, M., Biggin, M.D., Venters, B.J., Jiang, C., Li, J., Pugh, B.F., and Gilmour, D.S. (2008). NELF and GAGA factor are linked to promoter-proximal pausing at many genes in Drosophila. Mol. Cell. Biol. *28*, 3290–3300.

Lee, H., Kraus, K.W., Wolfner, M.F., and Lis, J.T. (1992). DNA sequence requirements for generating paused polymerase at the start of hsp70. Genes & Development *6*, 284–295.

Lee, M.G., Villa, R., Trojer, P., Norman, J., Yan, K.-P., Reinberg, D., Di Croce, L., and Shiekhattar, R. (2007). Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. Science *318*, 447–450.

Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. Annu. Rev. Genet. *34*, 77–137.

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H., and Kim, V.N. (2004). MicroRNA genes are transcribed by RNA polymerase II. The EMBO Journal *23*, 4051–4060.

Leibovitch, B.A., Schejter, E.D., Lu, Q., Wieschaus, E., Benjamin, L.R., Liu, Y., Gilmour, D.S., and Elgin, S.C.R. (2002). GAGA factor and the TFIID complex collaborate in generating an open chromatin structure at the Drosophila melanogaster hsp26 promoter. Mol. Cell. Biol. *22*, 6148–6157.

Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat. Rev. Genet. *13*, 233–245.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565–570.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707–719.

Liang, H.-L., Nien, C.-Y., Liu, H.-Y., Metzstein, M.M., Kirov, N., and Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature *456*, 400–403.

Lien, W.-H., Guo, X., Polak, L., Lawton, L., Young, R.A., Zheng, D., and Fuchs, E. (2011). Genome-wide maps of histone modifications unwind in vivo chromatin states of the hair follicle lineage. Cell Stem Cell *9*, 219–232.

Lim, C.Y. (2004). The MTE, a new core promoter element for transcription by RNA polymerase II. Genes & Development *18*, 1606–1617.

Lindeman, L.C., Andersen, I.S., Reiner, A.H., Li, N., Aanes, H., Østrup, O., Winata, C., Mathavan, S., Müller, F., Aleström, P., et al. (2011). Prepatterning of Developmental Gene Expression by Modified Histones before Zygotic Genome Activation. Developmental Cell 1–12.

Liu, Z., Scannell, D.R., Eisen, M.B., and Tjian, R. (2011). Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. Cell *146*, 720–731.

Loncar, D., and Singer, S.J. (1995). Cell membrane formation during the cellularization of the syncytial blastoderm of Drosophila. Proc Natl Acad Sci USA *92*, 2199–2203.

Lott, S.E., Villalta, J.E., Schroth, G.P., Luo, S., Tonkin, L.A., and Eisen, M.B. (2011). Noncanonical Compensation of Zygotic X Transcription in Early Drosophila melanogaster Development Revealed through Single-Embryo RNA-Seq. PLoS Biol *9*, e1000590.

Lowenhaupt, K., Cartwright, I.L., Keene, M.A., Zimmerman, J.L., and Elgin, S.C. (1983). Chromatin structure in pre- and postblastula embryos of Drosophila. Dev. Biol. *99*, 194–201.

Lu, X., Li, J.M., Elemento, O., Tavazoie, S., and Wieschaus, E.F. (2009). Coupling of zygotic transcription to mitotic control at the Drosophila mid-blastula transition. Development *136*, 2101–2110.

MacAlpine, D.M., Rodríguez, H.K., and Bell, S.P. (2004). Coordination of replication and transcription along a Drosophila chromosome. Genes & Development *18*, 3094–3105.

MacAlpine, H.K., Gordan, R., Powell, S.K., Hartemink, A.J., and MacAlpine, D.M. (2010). Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. Genome Research *20*, 201–211.

Magnani, L., Eeckhoute, J., and Lupien, M. (2011). Pioneer factors: directing transcriptional regulators within the chromatin environment. Trends in Genetics.

Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat. Rev. Genet. *11*, 761–772.

Martindale, M.Q. (2005). The evolution of metazoan axial properties. Nature Publishing Group *6*, 917–927.

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C., et al. (2008). Nucleosome organization in the Drosophila genome. Nature *453*, 358–362.

McDonald, K.L., Sharp, D.J., and Rickoll, W. (2012). Transmission electron microscopy of thin sections of Drosophila: conventional chemical fixation of embryos using trialdehyde. Cold Spring Harb Protoc *2012*, 516–520.

McKnight, S.L., and Miller, O.L. (1976). Ultrastructural patterns of RNA synthesis during early embryogenesis of Drosophila melanogaster. Cell *8*, 305–319.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Min, I.M., Waterfall, J.J., Core, L.J., Munroe, R.J., Schimenti, J., and Lis, J.T. (2011). Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes & Development *25*, 742–754.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. Nat Genet *37*, 1090–1097.

Montanuy, I., Hendrix, D.A., Torremocha, R., Hong, J.-W., Hernández-Munain, C., Zeitlinger, J., Suñé, C., Rokhsar, D.S., and Levine, M.S. (2008). Promoter influences transcription elongation: TATA-box element mediates the assembly of processive transcription complexes responsive to cyclin-dependent kinase 9. J. Biol. Chem. *283*, 7368–7378.

Morisato, D., and Anderson, K.V. (1995). Signaling pathways that establish the dorsal-ventral pattern of the Drosophila embryo. Annu. Rev. Genet. *29*, 371–399.

Moshkin, Y.M., Chalkley, G.E., Kan, T.W., Reddy, B.A., Ozgur, Z., van Ijcken, W.F.J., Dekkers, D.H.W., Demmers, J.A., Travers, A.A., and Verrijzer, C.P. (2012). Remodelers Organize Cellular Chromatin by Counteracting Intrinsic Histone-DNA Sequence Preferences in a Class-Specific Manner. Mol. Cell. Biol. *32*, 675–688.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nat Genet *39*, 1507–1511.

Müller, F., Lakatos, L., Dantonel, J., Strähle, U., and Tora, L. (2001). TBP is not universally required for zygotic RNA polymerase II transcription in zebrafish. Current Biology *11*, 282–287.

Nakakura, N., Miura, T., Yamana, K., Ito, A., and Shiokawa, K. (1987). Synthesis of heterogeneous mRNA-like RNA and low-molecular-weight RNA before the midblastula transition in embryos of Xenopus laevis. Dev. Biol. *123*, 421–429.

Narita, T., Yamaguchi, Y., Yano, K., Sugimoto, S., Chanarat, S., Wada, T., Kim, D.-K., Hasegawa, J., Omori, M., Inukai, N., et al. (2003). Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. Mol. Cell. Biol. *23*, 1863–1873.

Näär, A.M., Lemon, B.D., and Tjian, R. (2001). Transcriptional coactivator complexes. Annu. Rev. Biochem. *70*, 475–501.

Newport, J., and Kirschner, M. (1982a). A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. Cell *30*, 675–686.

Newport, J., and Kirschner, M. (1982b). A major developmental transition in early Xenopus embryos: II. Control of the onset of transcription. Cell *30*, 687–696.

Nègre, N., Brown, C.D., Ma, L., Bristow, C.A., Miller, S.W., Wagner, U., Kheradpour, P., Eaton, M.L., Loriaux, P., Sealfon, R., et al. (2011). A cis-regulatory map of the Drosophila genome. Nature *471*, 527–531.

Ni, T., Corcoran, D.L., Rach, E.A., Song, S., Spana, E.P., Gao, Y., Ohler, U., and Zhu, J. (2010). A paired-end sequencing strategy to map the complex landscape of transcription initiation. Nat Meth *7*, 521–527.

Nüsslein-Volhard, C., Wieschaus, E., and Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle inDrosophila melanogaster. Development Genes and Evolution *193*, 267–282.

O'Dor, E., Beck, S.A., and Brock, H.W. (2006). Polycomb group mutants exhibit mitotic defects in syncytial cell cycles of Drosophila embryos. Dev. Biol. *290*, 312–322.

Ohler, U., and Wassarman, D.A. (2010). Promoting developmental transcription. Development *137*, 15–26.

Ohler, U., Liao, G.-C., Niemann, H., and Rubin, G.M. (2002). Computational analysis of core promoters in the Drosophila genome. Genome Biol. *3*, RESEARCH0087.

Olins, D.E., and Olins, A.L. (2003). Chromatin history: our view from the bridge. Nat Rev Mol Cell Biol *4*, 809–814.

Ong, C.-T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation

of tissue-specific gene expression. Nature Publishing Group 1–11.

Pagans, S., Ortiz-Lombardía, M., Espinás, M.L., Bernués, J., and Azorín, F. (2002). The Drosophila transcription factor tramtrack (TTK) interacts with Trithorax-like (GAGA) and represses GAGA-mediated activation. Nucleic Acids Research *30*, 4406–4413.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., and Thomson, J.A. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell *1*, 299–312.

Papp, B., and Müller, J. (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes & Development *20*, 2041–2054.

Parkhurst, S.M., and Ish-Horowicz, D. (1991). wimp, a dominant maternal-effect mutation, reduces transcription of a specific subset of segmentation genes in Drosophila. Genes & Development *5*, 341–357.

Peel, A.D., Chipman, A.D., and Akam, M. (2005). Arthropod Segmentation: beyond the Drosophila paradigm. Nature Publishing Group *6*, 905–916.

Petruk, S., Tautz, D., Sedkov, Y., Lehmann, R., Johnston, D.M., rch, H.S.U., Hodgson, J.W., Schuh, R., Black, K.L., Seifert, E., et al. (2012). TrxG and PcG proteins but not methylated histones remain associated with DNA through replication. Cell *150*, 922–933.

Phillips, M.D., and Shearn, A. (1990). Mutations in *Polycombeotic,* A Drosophila Polycomb-Group Gene, Cause a Wide Range of Maternal and Zygotic Phenotypes. Genetics *125*, 91–101.

Pierce, B.A. (2009). Genetics Essentials: Concepts and Connections (W. H. Freeman).

Pirrotta, V. (1990). Transvection and long-distance gene regulation. Bioessays *12*, 409–414.

Prioleau, M.N., Huet, J., Sentenac, A., and Méchali, M. (1994). Competition between chromatin and transcription complex assembly regulates gene expression during early development. Cell *77*, 439–449.

Pritchard, D.K., and Schubiger, G. (1996). Activation of transcription in Drosophila embryos is a gradual process mediated by the nucleocytoplasmic ratio. Genes & Development *10*, 1131–1142.

Probst, A.V., Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. Nat Rev Mol Cell Biol *10*, 192–206.

Rach, E.A., Ingham, P.W., Yuan, H.-Y., Howard, K.R., Majoros, W.H., Ish-Horowicz, D., Tomancak, P., and Ohler, U. (2009). Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the Drosophila genome. Genome Biol. *10*, R73.

Rach, E.A., Winter, D.R., Benjamin, A.M., Corcoran, D.L., Ni, T., Zhu, J., and Ohler, U. (2011). Transcription Initiation Patterns Indicate Divergent Strategies for Gene Regulation

at the Chromatin Level. PLoS Genet *7*, e1001274.

Rasmussen, E.B., and Lis, J.T. (1993). In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc Natl Acad Sci USA *90*, 7923–7927.

Rivera-Pomar, R., and Jäckle, H. (1996). From gradients to stripes in Drosophila embryogenesis: filling in the gaps. Trends Genet. *12*, 478–483.

Ross, J.F. (2001). Active repression and E2F inhibition by pRB are biochemically distinguishable. Genes & Development *15*, 392–397.

Roth, S., and Lynch, J.A. (2009). Symmetry Breaking During Drosophila Oogenesis. Cold Spring Harbor Perspectives in Biology *1*, a001891–a001891.

Rothe, M., Pehl, M., Taubert, H., and Jäckle, H. (1992). Loss of gene function through rapid mitotic cycles in the Drosophila embryo. Nature *359*, 156–159.

Rouget, C., Papin, C., Boureux, A., Meunier, A.-C., Franco, B., Robine, N., Lai, E.C., Pelisson, A., and Simonelig, M. (2010). Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. Nature *467*, 1128–1132.

Rougvie, A.E., and Lis, J.T. (1988). The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. Cell *54*, 795–804.

Rubin, G.M. (2000). A brief history of Drosophila's contributions to genome research. Science *287*, 2216–2218.

Sanchez-Elsner, T., Gou, D., Kremmer, E., and Sauer, F. (2006). Noncoding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax. Science *311*, 1118–1123.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.-Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. Cell *132*, 887–898.

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009). Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. PLoS Biol *7*, e13.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nat. Rev. Genet. *8*, 9–22.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. Cell *148*, 458–472.

Shermoen, A.W., and O'Farrell, P.H. (1991). Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. Cell *67*, 303–310.

Simon, J.A., and Kingston, R.E. (2009). Mechanisms of Polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Biol 1–12.

Smale, S.T., and Baltimore, D. (1989). The "initiator" as a transcription control element.

Cell *57*, 103–113.

Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. Annu. Rev. Biochem. *72*, 449–479.

Smith, E., Lin, C., and Shilatifard, A. (2011). The super elongation complex (SEC) and MLL in development and disease. Genes & Development *25*, 661–672.

Smith, E.R., Lee, M.G., Winter, B., Droz, N.M., Eissenberg, J.C., Shiekhattar, R., and Shilatifard, A. (2008). Drosophila UTX is a histone H3 Lys27 demethylase that colocalizes with the elongating form of RNA polymerase II. Mol. Cell. Biol. *28*, 1041–1046.

Sokac, A.M., and Wieschaus, E. (2008). Local actin-dependent endocytosis is zygotically controlled to initiate Drosophila cellularization. Developmental Cell *14*, 775–786.

Spiker, S., Murray, M.G., and Thompson, W.F. (1983). DNase I sensitivity of transcriptionally active genes in intact nuclei and isolated chromatin of plants. Proc Natl Acad Sci USA *80*, 815–819.

St Johnston, R.D., and Gelbart, W.M. (1987). Decapentaplegic transcripts are localized along the dorsal-ventral axis of the Drosophila embryo. The EMBO Journal *6*, 2785–2791.

Stark, A., Lin, M.F., Kheradpour, P., Pedersen, J.S., Parts, L., Carlson, J.W., Crosby, M.A., Rasmussen, M.D., Roy, S., Deoras, A.N., et al. (2007). Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. Nature *450*, 219–232.

Sullivan, W., and Theurkauf, W.E. (1995). The cytoskeleton and morphogenesis of the early Drosophila embryo. Current Opinion in Cell Biology *7*, 18–22.

Svejstrup, J.Q. (2012). Transcription: another mark in the tail. The EMBO Journal *31*, 2753–2754.

Tadros, W., and Lipshitz, H.D. (2009). The maternal-to-zygotic transition: a play in two acts. Development *136*, 3033–3042.

Tadros, W., Goldman, A.L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T.R., Westwood, J.T., Smibert, C.A., and Lipshitz, H.D. (2007). SMAUG is a major regulator of maternal mRNA destabilization in Drosophila and its translation is activated by the PAN GU kinase. Developmental Cell *12*, 143–155.

The modENCODE Consortium, Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., et al. (2010). Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. Science *330*, 1787–1797.

Theurkauf, W.E., and Hawley, R.S. (1992). Meiotic spindle assembly in Drosophila females: behavior of nonexchange chromosomes and the effects of mutations in the nod kinesin-like protein. J. Cell Biol. *116*, 1167–1180.

Thomas, M.C., and Chiang, C.-M. (2006). The General Transcription Machinery and General Cofactors. Critical Reviews in Biochemistry and Molecular Biology *41*, 105–178.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S.E., Richards, S.,

Ashburner, M., Hartenstein, V., Celniker, S.E., et al. (2002). Systematic determination of patterns of gene expression during Drosophila embryogenesis. Genome Biol. *3*, RESEARCH0088.

Tomancak, P., Berman, B.P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S.E., and Rubin, G.M. (2007). Global analysis of patterns of gene expression during Drosophila embryogenesis. Genome Biol. *8*, R145.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Tursun, B. (2012). Cellular reprogramming processes in Drosophila and C. elegans. Current Opinion in Genetics & Development *22*, 475–484.

van Ingen, H., van Schaik, F.M.A., Wienk, H., Ballering, J., Rehmann, H., Dechesne, A.C., Kruijzer, J.A.W., Liskamp, R.M.J., Timmers, H.T.M., and Boelens, R. (2008). Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. Structure/Folding and Design *16*, 1245–1256.

Vastenhouw, N.L., and Schier, A.F. (2012). Bivalent histone modifications in early embryogenesis. Current Opinion in Cell Biology *24*, 374–386.

Vastenhouw, N.L., Zhang, Y., Woods, I.G., Imam, F., Regev, A., Liu, X.S., Rinn, J., and Schier, A.F. (2010). Chromatin signature of embryonic pluripotency is established during genome activation. Nature *464*, 922–926.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W.M.P., van Schaik, F.M.A., Varier, R.A., Baltissen, M.P.A., Stunnenberg, H.G., Mann, M., and Timmers, H.T.M. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell *131*, 58–69.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., et al. (1998a). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes & Development *12*, 343–356.

Wada, T., Takagi, T., Yamaguchi, Y., Watanabe, D., and Handa, H. (1998b). Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. The EMBO Journal *17*, 7395–7403.

Waddington, C.H. (1957). The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser. ix+–262.

Wakayama, T., Perry, A.C., Zuccotti, M., Johnson, K.R., and Yanagimachi, R. (1998). Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. Nature *394*, 369–374.

Wang, X., Hang, S., Prazak, L., and Gergen, J.P. (2010). NELF potentiates gene transcription in the Drosophila embryo. PLoS ONE *5*, e11498.

Wang, X., Lee, C., Gilmour, D.S., and Gergen, J.P. (2007). Transcription elongation controls cell fate specification in the Drosophila embryo. Genes & Development *21*, 1031–1036.

Weake, V.M., and Workman, J.L. (2011). SAGA function in tissue-specific gene expression. Trends Cell Biol.

West, M.L., and Corden, J.L. (1995). Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. Genetics *140*, 1223.

White, R.J. (2008). RNA polymerases I and III, non-coding RNAs and cancer. Trends in Genetics *24*, 622–629.

Wieschaus, E., Nüsslein-Volhard, C., and Jürgens, G. (1984). Mutations affecting the pattern of the larval cuticle inDrosophila melanogaster. Development Genes and Evolution *193*, 296–307.

Wilkins, R.C., and Lis, J.T. (1997). Dynamics of potentiation and activation: GAGA factor and its role in heat shock gene regulation.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H. (1997). Viable offspring derived from fetal and adult mammalian cells. Nature *385*, 810–813.

Wu, S.-F., Zhang, H., and Cairns, B.R. (2011). Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm. Genome Research *21*, 578–589.

Xie, F., and Orr-Weaver, T.L. (2008). Isolation of a Drosophila amplification origin developmentally activated by transcription. Proc Natl Acad Sci USA *105*, 9651–9656.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell *97*, 41–51.

Yean, D., Erickson, J.W., Gralla, J., and Cline, T.W. (1997). Transcription reinitiation rate: a special role for the TATA box. Mol. Cell. Biol. *17*, 3809–3816.

Yean, D., Yean, D., Gralla, J.D., and Gralla, J.D. (1999). Transcription reinitiation rate: a potential role for TATA box stabilization of the TFIID:TFIIA:DNA complex. Nucleic Acids Research *27*, 831–838.

Yuan, G.C. (2005). Genome-Scale Identification of Nucleosome Positions in S. cerevisiae. Science *309*, 626–630.

Zalokar, M. (1976). Autoradiographic study of protein and RNA formation during early development of Drosophila eggs. Dev. Biol. *49*, 425–437.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes & Development *25*, 2227–2241.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.-W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet *39*, 1512–1516.

Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biol *15*, 1263–1271.

Zhang, M., Wang, F., Kou, Z., Zhang, Y., and Gao, S. (2009a). Defective chromatin structure in somatic cell cloned mouse embryos. Journal of Biological Chemistry *284*, 24981–24987.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhang, Y., Moqtaderi, Z., Rattner, B.P., Euskirchen, G., Snyder, M., Kadonaga, J.T., Liu, X.S., and Struhl, K. (2009b). Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. Nature Structural & Molecular Biology *16*, 847–852.

Zhang, Z., Wippo, C.J., Wal, M., Ward, E., Korber, P., and Pugh, B.F. (2011). A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. Science *332*, 977–980.

Zhao, X.D., Han, X., Chew, J.L., Liu, J., Chiu, K.P., Choo, A., Orlov, Y.L., Sung, W.-K., Shahab, A., Kuznetsov, V.A., et al. (2007). Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. Cell Stem Cell *1*, 286–298.